Title
Development of X-ray Methods for the Investigation of Protein Dynamics

Author
Wolff, Alexander Michael

Publication Date
2020

Peer reviewed|Thesis/dissertation

Development of X-ray Methods for the Investigation of Protein Dynamics

by
Alexander Michael Wolff

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
James S. Fraser
EF2645E0B04B452...

James S. Fraser
Chair

DocuSigned by:
Andrej Sali
[redacted]

Andrej Sali

DocuSigned by:
Natalia Jura
E438674A382B42F...

Natalia Jura

Committee Members

Acknowledgments

I would like to begin by thanking James Fraser. James's enthusiasm for science and his willingness to venture off the beaten path were key factors in my decision to pursue my PhD at UCSF. Throughout my time in the lab, James has shown me the value of clear communication, while training me to ask the big questions. His optimism and trust encouraged me to explore science broadly, while forming a strong network of collaborators. These factors brought me joy during my graduate work.

UCSF is home to a supportive and collaborative group of professors, many of whom I owe a debt of gratitude. Early on in my graduate studies, I received excellent mentorship and guidance from David Agard and Martin Kampmann. As I progressed, Andrej Sali and Natalia Jura served as supportive committee members.

Without my mentors at the University of Wyoming, I never would have made the journey to UCSF. Mark Stayton provided immeasurable guidance and wisdom during my early years as a scientist. Dan Levy, Jay Gatlin, Don Jarvis, Franco Basile, Marci Smith, and countless others provided the encouragement I needed to push myself.

During my time at UCSF, I have had the opportunity to work closely with many collaborators, but I am especially grateful to Nick Sauter and Michael Wall for their time. Nick and his group provided patient guidance on the analysis of X-ray scattering data. Michael served as a mentor throughout most of my graduate career and was gracious with his time and energy. He provided much-needed counsel on the analysis of diffuse scattering data, and was a welcoming host when I visited Los Alamos.

I want to thank everyone in the Fraser Lab for their support and encouragement over the past few years. David Mavor reminded me to be joyful and light-hearted. Ben Barad took me under his wing, helping me to get up to speed on coding and coffee while setting a high standard for professionalism. Justin Biel taught me about molecular visualization while demonstrating the value of being meticulous and thoughtful. John Lee and I bonded as burrito

brothers, with help from Roberto Chica. John's ability to relate and to listen when times were tough helped me persevere. John and Jen Michaud-Lee opened their hearts and their home countless times, and I value the wonderful moments I've spent with them. I am also indebted to Iris Young, who lent an ear and a hand at key moments in my life – what a gift! Finally, I am fortunate to have worked with Michael Thompson throughout my graduate career. The teamwork we fostered and the dreams we chased were the highlights of my scientific work so far. Mike's vision and guidance, as well as his persistent motivation, have kept me on track during my graduate studies.

Classmates and friends have provided the levity I needed at crucial moments in my graduate career. I am very thankful to Tamas Nagy and Andrea Eastes, with whom I have shared countless mountain adventures. Arthur Melo, David Baskin, and Vikram Ramkumar have served as an enthusiastic climbing partners, keeping positivity alive regardless of outcomes. Daniel Sullivan has lent an empathetic ear and pushed me to grow as a person.

Lastly, I owe more than I will ever know to my family. The Luhrs provided positive encouragement, and I am especially grateful for the interest Dan Luhr took in my scientific endeavors over the years. Leslie, I am grateful for the support you provided along the way. I am also grateful for the time I had with my two kitties, Dulcinea and Saros. You furry sages reminded me that living in the moment and playing is as good as life gets. Grandpa Wolff, thanks for teaching me how to dream, to take meticulous care in my work, and to raise my awareness. You were the first scholar I knew. Grandma Wolff, you taught me to be pragmatic, grounded, and to empathize with others who might appear different from myself. Dad, you taught me to be irreverent and to take my own path. Mom, you remind me to remain aware of what matters most in life. Without these guiding forces, my feet would have wandered along another path, so I am filled with gratitude for each and every one of you.

Contributions

Chapter I

Wolff, A. M., Young, I. D., Sierra, R. G., Brewster, A. S., Martynowycz, M. W., Nango, E., Sugahara, M., Nakane, T., Ito, K., Aquila, A., Bhowmick, A., Biel, J. T., Carbajo, S., Cohen, A. E., Cortez, S., Gonzalez, A., Hino, T., Im, D., Koralek, J. D., … Thompson, M. C. (2020). Comparing serial X-ray crystallography and microcrystal electron diffraction (MicroED) as methods for routine structure determination from small macromolecular crystals. *IUCrJ*, *7*(2), 306–323. https://doi.org/10.1107/S205225252000072X

Chapter II

Thompson, M. C., Barad, B. A., Wolff, A. M., Sun Cho, H., Schotte, F., Schwarz, D. M. C., Anfinrud, P., & Fraser, J. S. (2019). Temperature-jump solution X-ray scattering reveals distinct motions in a dynamic enzyme. *Nature Chemistry*, *11*(11), 1058–1066. https://doi.org/10.1038/s41557-019-0329-3

Chapter IV

Wall, M. E., Wolff, A. M., & Fraser, J. S. (2018). Bringing diffuse X-ray scattering into focus. *Current Opinion in Structural Biology*, *50*, 109–116. https://doi.org/10.1016/j.sbi.2018.01.009

Development of X-ray Methods for the Investigation of Protein Dynamics

Alexander Michael Wolff

Abstract

Throughout history, methodological innovations have resulted in breakthroughs in our understanding of biology. Methods for determining static protein structures, as well as those for probing protein dynamics, are well-established. Nonetheless, visualizing molecules as dynamic entities that respond to their environment is still an outstanding challenge. Specifically, it is challenging to measure the spatial position of all the atoms within a molecule as a function of time. That challenge is the broad focus of this dissertation.

In chapter one, I begin by diving into modern crystallographic techniques that enable one to solve protein structures from sub-micron-sized crystals. I compare and contrast two methods, serial crystallography and electron crystallography, asking how each technique affects the protein's structure. A primary factor differentiating these two methods is the temperature of the sample during the experiment. Despite this difference, both methods enable one to solve high-resolution structures from small crystals. This is advantageous for time-resolved experiments. Since there are fewer molecules in a small crystal, the perturbation is more uniform, which provides a clearer time-resolved signal.

In chapter two, I investigate temperature-jumps as a generalized perturbation for resolving the energy landscape of proteins. In this work, I focus on solution scattering experiments, which allow one to examine large-scale perturbations to a protein, as well as changes in the solvent shell surrounding the molecule. By mutating selected residues, we inhibited specific protein motions. Comparing these mutants to the wild-type protein allowed us to resolve the motions driven by an infrared laser. Nonetheless, we wished to gain all-atom spatial resolution, which required us to perform a temperature-jump within the context of crystallography rather than solution scattering.

In chapter three, I expand upon the temperature-jump detection method described in chapter two. By adapting this method to accommodate X-ray diffraction images, I demonstrate that we can detect temperature-jumps within a crystalline context. This is a crucial step in the development of a generalized perturbation for time-resolved crystallography. Given the timescale of the measurements, reading out the temperature directly from the X-ray data is the only effective way to track the sample's response. Thus, our method offers proof-of-principle that IR laser-based temperature-jumps are feasible for time-resolved crystallography. While measuring the diffuse scattering signal is useful for temperature-jump detection, the diffuse signal also holds the potential to inform our understanding of protein dynamics.

In chapter four, I review the field of macromolecular diffuse scattering, as of late 2017. I begin by considering data collection practices, which requires extremely careful and controlled measurements. Then I examine different group's approaches to processing the data, as well as their models of the disorder that drives it. Finally, I consider the broader impact of diffuse analysis upon the field, ranging from the improvement of molecular dynamics forcefields to improved phasing and resolution extension. While these impacts hold exciting implications, it is clear that collecting high-quality is the first challenge to solve.

In chapter five, I examine the challenges of collecting high-quality diffuse scattering from protein crystals. I describe how parasitic scattering can confound our ability to develop rigorous models of the crystalline disorder that gives rise to the diffuse signal. Then I work through experimental measures that we took to minimize parasitic scattering while maximizing diffuse scatter driven by protein motions.

Table of Contents

List of Figures

Chapter I

Chapter II

List of Tables

Chapter I

Chapter II

Chapter III

Chapter I

Comparing serial X-ray crystallography and microcrystal electron diffraction (MicroED) as methods for routine structure determination from small macromolecular crystals

Contributing Authors

Alexander M Wolff [1,2], Iris D Young [2,3], Raymond G Sierra [4], Aaron S Brewster [3], Michael W Martynowycz [5,6] Eriko Nango [7,8], Michihiro Sugahara [7], Takanori Nakane [9], Kazutaka Ito [2,10], Andrew Aquila [4], Asmit Bhowmik [3], Justin T Biel [1,2], Sergio Carbajo [4], Aina E Cohen [11]; Saul Cortez [12], Ana Gonzalez [11], Tomoya Hino [13,14], Dohyun Im [8], Jake D Koralek [4], Minoru Kubo [7,15], Tomas S Lazarou [16], Takashi Nomura [7], Shigeki Owada [7,22], Avi J Samelson [17], Tomoyuki Tanaka [7,8], Rie Tanaka [7,8], Erin M Thompson [2,18], Henry van den Bedem [19], Rahel A Woldeyes [2,18], Fumiaki Yumoto [20], Wei Zhao [21], Kensuke Tono [7,22], Sébastien Boutet [4], So Iwata [7,8], Tamir Gonen [5,6,23], Nicholas K Sauter [3], James S Fraser [2], Michael C Thompson [2]

1. Graduate Program in Biophysics, University of California, San Francisco, San Francisco, CA, USA
2. Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA
3. Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
4. Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA, USA
5. Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA, USA
6. Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA
7. RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo, 679-5148, Japan
8. Department of Cell Biology, Graduate School of Medicine, Kyoto University, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
9. Department of Biological Science, Graduate School of Science, The University of Tokyo, Tokyo, Japan
10. Laboratory for Drug Discovery, Pharmaceuticals Research Center, Asahi Kasei Pharma Corporation, 632-1 Mifuku, Izunokuni-shi, Shizuoka 410-2321, Japan.

11. SSRL, SLAC National Accelerator Laboratory, Menlo Park, CA, USA
12. Department of Biology, San Francisco State University, San Francisco, CA, USA
13. Department of Chemistry and Biotechnology, Graduate School of Engineering, Tottori University, 4-101, Koyama-cho Minami, Tottori 680-8552, Japan
14. Center for Research on Green Sustainable Chemistry, Tottori University, Tottori, Japan
15. Graduate School of Life Science, University of Hyogo, Ako-gun, Hyogo 678-1297, Japan
16. Department of Chemistry, New York University, New York, NY, USA
17. Institute for Neurodegenerative Diseases, University of California, San Francisco, San Francisco, CA, USA
18. Graduate Program in Chemistry and Chemical Biology, University of California, San Francisco, San Francisco, CA, USA
19. Bioscience Department, SLAC National Accelerator Laboratory, Menlo Park, CA, USA
20. Structural Biology Research Center, Institute of Materials Structure Science, KEK/High Energy Accelerator Research Organization, Tsukuba, Ibaraki 305-0034, Japan
21. Department of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA
22. Japan Synchrotron Radiation Research Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5198, Japan
23. Department of Physiology, University of California, Los Angeles, Los Angeles, CA

Preface

The bulk of this work appears as Wolff *et al.*, published in *IUCrJ* in 2020.

This work evolved from my earliest endeavors in the Fraser Lab where I joined Michael Thompson to develop time-resolved temperature-jump diffraction experiments. We planned to use serial crystallography methods, which require large quantities of microcrystals. At that time, serial crystallography (SFX) was at the frontier of structural biology. Thus, I created slurries of microcrystals, and we tested their suitability for SFX experiments. As the field of structural biology expanded, it became clear that we could also use micro electron diffraction (microED) to investigate our protein's structure. Since SFX experiments occur at room temperature, while microED experiments occur at cryogenic temperatures, we wondered whether the resulting structures would differ. The crystals utilized for microED experiments are so small that we hypothesized that we could freeze them fast enough to lock the room temperature structure in place. While conducting the experiment, our microcrystals were still too large for microED. Nonetheless, with the help of our collaborators we were able to mill the crystals down in size, and then solve a structure using microED. Thus, our comparison of microED and SFX, initially

an offshoot of our temperature-jump work, evolved into the independent analysis presented in this chapter.

Abstract

Innovative methods are facilitating structural studies from ever smaller crystals, yet targeted creation and optimal data collection from microcrystals are still open challenges. Moreover, the effect of sample preparation for various serial and microcrystal electron diffraction (MicroED) methods on protein conformation has not been examined. Using Cyclophilin A (CypA) as a model system, this work reveals that structures solved using serial XFEL methods differ from a structure solved from focused ion-beam (FIB)-milled crystals using MicroED. While serial XFEL structures resemble published structures from room temperature single crystal rotational experiments, a MicroED structure more closely resembles published work from cryogenic single crystal experiments. Additionally, unit cell dimensions varied as a function of delivery method for serial XFEL datasets. These results indicate that crystals are affected by delivery strategy, and that protein structure could be affected for molecules where crystal contacts modulate conformation. Overall, this work provides a route for microcrystal optimization, reveals that sample delivery methods for serial XFEL experiments do not perturb CypA's structure, and demonstrates that sample temperature may provide a critical differentiation between microcrystal structures solved using serial XFEL and those solved using MicroED.

Introduction

In macromolecular crystallography, collecting full datasets from small crystals has been a challenge, because their weaker diffracting power limits the amount of signal that can be successfully obtained before the effects of X-ray radiation damage become significant. Thus, crystallographers have always been faced with a practical need to either optimize the growth of

relatively large crystals, or to make the most of smaller crystals by implementing clever data collection and merging strategies[1,2]. Methodological advances that facilitated measurement of diffraction data from smaller and smaller crystals, such as the introduction of crystal cryocooling and the development of microfocus X-ray beams, have enabled structure determination of increasingly challenging targets, for which large crystals could only be obtained with difficulty, if at all[3,4]. Additionally, small crystals have proven advantageous in a number of other contexts. For example, when crystals display long-range disorder, such as high mosaicity, a reduction in the total number of mosaic blocks reduces the spread of Bragg peaks and generally improves the overall data quality[5]. Small crystals are also advantageous when the diffraction experiment is preceded by a perturbation to the crystal. This includes common crystal treatments such as flash cooling or ligand soaking, as well as more uncommon perturbations, like stimulation of crystallized molecules for time-resolved experiments[6,7]. Because they have substantially less volume and a limited number of unit cells, perturbations can be applied more rapidly and uniformly to smaller crystals than to larger ones, and smaller crystals accumulate less strain resulting from changes in crystal lattice dimensions. The development of protein "microcrystallography" techniques, which are optimized for measuring crystals whose physical dimensions are tens of microns or smaller, has offered access to these opportunities and benefits, and led to a shift in what is considered to be a valuable specimen for experimental characterization.

The past decade has seen an explosion of new technology for protein microcrystallography. The increased brightness available for crystallography at modern X-ray light sources, including synchrotrons and X-ray free-electron lasers (XFELs), led to the development of "serial crystallography." In a serial crystallography experiment, the X-ray beam is typically very bright and tightly focused, so that extremely short exposure times produce measurable diffraction images, even for very small crystals[8]. The intense X-ray beams used for these experiments destroy the samples rapidly, allowing only a single diffraction image to be

collected per crystal. Therefore, the crystals must be rapidly (or “serially”) replenished at the X-ray interaction point in order for the experiment to be efficient. By measuring single diffraction snapshots of many randomly oriented crystals, it is possible to completely sample the reflections in reciprocal space and integrate the Bragg intensities. Importantly, serial crystallography experiments are generally conducted at room temperature, potentially giving more physiologically-relevant insight into molecular structure by avoiding the artifacts associated with cryocooling.

Alongside the development of serial crystallography, major recent breakthroughs have been made in the field of microcrystal electron diffraction (MicroED). Specifically, advancements have resulted in the facile collection of continuous rotation datasets[9,10] from flash-cooled microcrystals using a transmission electron microscope that is equipped with a compustage and cryoholder, and operating in diffraction mode. Because the microscopes required for MicroED are now widespread as a result of booming interest in electron cryomicroscopy (cryoEM), MicroED holds great potential for the determination of both protein and small molecule structures[11]. Collectively, these new frontiers in macromolecular microcrystallography have created new opportunities for structural biology. Examples include structure determination from crystals as small as a few hundred nanometers in each of their dimensions[8,11,12], and a new generation of challenging time-resolved measurements[13,14] at high spatial and temporal resolution.

Despite the interesting possibilities that are now within reach, the optimization of sample preparation and data collection protocols for microcrystallography experiments remains challenging. First, it is necessary to decide which measurement technique (i.e. X-ray or MicroED) is best suited to a given sample or research question. Then, if appropriately-sized crystals are not obtained serendipitously, the experimenter must generate microcrystals with the correct size and density (crystals / uL), either by targeted growth or by manipulation of larger crystal specimens. Next, it is essential to choose an appropriate method for delivering the

microcrystals to the X-ray beam or to the column of the electron microscope. In the case of serial X-ray crystallography, multiple strategies have been explored for rapidly replenishing crystals at the X-ray interaction point. In addition to fixed-target approaches[15–18], where crystals are mounted on a solid support and moved through the X-ray interaction region using automation, several methods have come into widespread use that exploit microfluidics to create free-standing streams, or "jets," of microcrystal slurries. Various different types of microfluidic devices, collectively referred to as "sample injectors," have been developed for this purpose[19,20]. Each type uses a different physical principle for carrying microcrystals to the X-ray beam by generating a stream of flowing liquid that is tens to hundreds of microns in diameter, and each method subjects the crystals to different conditions which could potentially affect the quality of data acquired or the structure of the molecule itself. These conditions include exposure to strong electric fields[19], high pressure[20], and additives[21] that change the chemical properties of mother liquor solutions.

Similarly in MicroED several sample preparation methods have been reported such as direct pipetting of nanocrystals onto EM grids[22]; sonication, vortexing, vigorous pipetting, or crushing to break big crystals into fragments and create a nanocrystal slurry[12]. Crystals are then drop cast onto EM grids, traditionally used for cryoEM, then the grids are blotted of excess solvent and flash frozen in supercooled ethane. Once frozen, microcrystals prepared in this fashion can be used directly for data collection, or they can be subjected to a milling procedure that utilizes a scanning electron microscope with a focused ion beam (FIB-SEM) to create crystalline lamellae of the desired thickness[23–25]. Because electrons interact with matter more strongly than X-rays do, the ideal crystal thickness for MicroED measurements is only several hundred nanometers[25] and the milling process is critical for samples that exceed this thickness. As for serial crystallography, sample preparation and delivery for MicroED involves subjecting crystals to unusual conditions that are not typically encountered when samples are prepared for traditional crystallographic experiments. These conditions include dehydration and exposure to

shear forces that are produced by the flow of solvent during blotting[24], as well as potential damage induced by FIB milling. All of these considerations create a complex landscape, and designing the best experiment for a new system of interest is often a non-trivial process. This calculus is further complicated by the fact that the extent to which the unusual experimental conditions affect the quality of data acquired, or the structure of the molecule itself, have not been rigorously characterized.

Here, we discuss the planning, optimizing, and executing protein microcrystallography experiments, using human cyclophilin A (CypA) as a model protein system. CypA is a proline isomerase enzyme that is highly abundant in human cells, and plays important biological roles as both a protein folding chaperone and modulator of intracellular signaling pathways. Prior work has shown that CypA readily forms large crystals, which have been successfully used for traditional rotation crystallography at synchrotron X-ray sources and for fixed target measurements at an XFEL source[26,27]. Starting from crystallization conditions that produce large (hundreds of microns in each dimension) CypA crystals, we optimized the preparation of high-density microcrystal slurries. We then used these samples for a variety of microcrystallography experiments, including serial X-ray crystallography with three different microfluidic sample injectors, and MicroED. Because the data collected across the different types of experiments were derived from similarly prepared microcrystal samples, and analyzed using comparable protocols, we were able to perform a rigorous comparison of the results. For each method, we evaluate the ease of sample preparation and delivery, the statistical quality of the measured data, and the properties of the resulting atomic models. Our results illustrate the inherent strengths and weaknesses of these new and exciting techniques for macromolecular microcrystallography, and lay out a roadmap for optimization of this promising category of experiments.

Methods

Protein Expression & Purification

Wild-type human Cyclophilin A (CypA) was expressed and purified as previously described[26]. Briefly, following purification, the protein was stored in a solution of 20 mM HEPES pH 7.5, 20 mM NaCl, 0.5 mM TCEP at 4° C until use. Finally, samples were concentrated using amicon centrifugal filters, then crystallized as described in section 2.2.

Crystal Formation & Optimization

For exploration of CypA's crystallization phase space, crystallization trays were set as follows. Well solutions containing 100 mM HEPES pH 7.5, 5 mM TCEP, and PEG 3350 (concentration varied) were distributed into Greiner 96-well Imp@ct microbatch crystallization plates. Each well contained 2 µL of the respective well solution, mixed with 2 µL of protein at the respective concentration. These drops were then vapor sealed using 12 µL of paraffin oil. For large-scale production of crystals in batch, 600 µL of protein at 60 mg/mL was combined with 400 µL of 50% PEG 3350 in a glass vial and stirred with a stir bar at a constant rate (RPM varied). Crystallization was robust over a temperature range spanning 20-25°C.

Crystal Analysis

Raw images of microcrystal slurries were taken under a Nikon Ti microscope in differential interference contrast mode, with a Nikon DS-Qi2 camera. Data were interpreted using Fiji software[28]. In addition to imaging crystalline slurries, particle densities (crystals/mL) were analyzed using an INCYTO C-Chip. Diffraction tests were carried out at Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2 using a 20 µm X 40 µm beam at 0.9795 Å. Crystals were diluted and loaded onto a Mitegen MicroMesh 700/25 Loop. Frames were collected for 1.0 s with a 1.0° oscillation. The angular extent of diffraction and unit cell dimensions were assessed using ADXV.

Sample Preparation for Serial X-ray Experiments

Crystals were formed in batch, as described in section 2.2, at a constant stir rate of 500

RPM. Further preparation was determined by delivery method. When using the MESH injector[19], the microcrystal slurry was delivered as is, in a Hamilton syringe. When using the LCP injector[20], crystals were mixed with viscogens, either PEO, LCP, or cellulose. For PEO mixtures, microcrystal slurry was combined with a viscogen consisting of 10% PEG + 10% PEO, and various ratios of crystal slurry to viscogen were tested. For LCP mixtures, crystal slurries were centrifuged, the supernatant was removed, with a minimal volume (100 µL) added back to suspend the crystals. The crystals were then mixed with Monoolein (9.9 MAG) in 1:1.5 mass-to-mass ratio using coupled glass syringes[29]. For cellulose mixtures, crystal slurries were centrifuged, the supernatant was removed, and crystals were directly mixed with a 20% hydroxyethyl cellulose in a 1:9 crystal-to-cellulose ratio, as previously described[21].

Serial X-ray Data Collection & Analysis

For the MESH & LCP XFEL datasets, we collected data at LCLS-MFX[30] on an MX170-HS Rayonix detector in 2-by-2 binning mode. Crystals were delivered to the XRD interaction point using either a MESH injector[19,31] or an LCP injector[20]. Data were collected with a 3 µm beam at 9.5 keV energy, pulsed at 10 Hz, with a pulse duration of 40 fs on average. Powder diffraction patterns of silver(I) behenate were used to estimate the detector distance. The *cctbx.xfel* GUI was used for real-time feedback on the hit-rate and indexing-rate, as well as to submit processing jobs onsite. Data were indexed and integrated using *dials.stills_process*[32]. Initial indexing results were used to refine the detector model, as well as crystal models[33]. Refinement of the detector distance and panel geometry improved the agreement between measured and predicted spots. Data were then merged and post-refined using *cxi.merge*[34]. Error estimates were treated according to the Ev11 method[33,35], wherein error estimates were increased using terms refined from the measurements until they could better explain the uncertainty observed in the merged reflection intensities. For the cellulose XFEL dataset, we collected data at SACLA[36] using a Diverse Application Platform for Hard X-ray Diffraction in SACLA (DAPHNIS)[37] at BL2[38]. Diffraction images were collected using a custom-built 4M pixel

detector with multi-port CCD (mpCCD) sensors[39]. Data collection was supported by a real-time data processing pipeline[40] developed on Cheetah[41]. Identified hit images were processed in CrystFEL version 0.6.3[42]. Diffraction spots were indexed by DirAx[43]. Intensities were merged by Monte Carlo integration with the *process_hkl* command in the CrystFEL suite with linear scale factors and per-image resolution cutoff. We note that the data collected at SACLA could not be processed using *dials.stills_process* due to spot shape irregularities that are an artifact of the mpCCD detector. Specifically, when strong X-ray signals are recorded on this detector, large streaks are produced across horizontal rows of pixels. Serial crystallography data analysis at SACLA relies on the CrystFEL pipeline, which has been optimized to handle the idiosyncrasies of the hardware at that facility.

MicroED Sample Preparation

Samples for MicroED were prepared as previously described[24]. A 2 µL aliquot of crystals from the batch solution was applied to a glow-discharged Quantifoil Cu200 mesh R2/2 holey carbon grid. The grid was gently blotted from the back (by hand) in an FEI Vitrobot for 10 s at 100% humidity and then vitrified in liquid ethane. Grids were stored in liquid nitrogen until further use, and all additional sample manipulation and data collection procedures were performed under cryogenic conditions. Prior to data collection, the grids were clipped and loaded into an FEI Versa FIB/SEM at liquid-nitrogen temperature and milled as previously described[24]. The grids were coated with a thin layer of amorphous platinum to increase the contrast during FIB/SEM imaging[25]. Large crystals (10-50 um) near the center of the grid square were identified using 2 kV SEM. Crystals were milled using a 30 kV gallium ion beam with a stepwise decreasing beam current as the sample slowly approached its final thickness of approximately 200 nm. The final 10 nm on either side of the crystalline lamellae were milled at 10 pA to polish the crystalline surface.

MicroED Data Collection & Analysis

Grids containing milled crystals were transferred into an FEI Arctica TEM operating at an

accelerating voltage of 200 kV under liquid nitrogen. Crystalline lamellae were identified initially in an all-grid atlas taken at 155x magnification, where crystals were apparent as semitransparent areas suspended over a sharp, straight strip of empty area created by the milling process. Continuous rotation MicroED data were collected in diffraction mode over an angular wedge between -60° and 0° from the untilted orientation at a rotation rate of 0.3° / s. The camera length was set to 2055 mm, and frames were read out every 2 seconds. Data were recorded on a CetaD detector operating in rolling shutter mode with 2-by-2 binning. The camera length was calibrated using a molybdenum foil. MicroED data were converted from the FEI SER format to SMV for data analysis using in-house software that is freely available (https://cryoem.ucla.edu/). Data were indexed and integrated with XDS and scaled in XSCALE[44].

Model Refinement and Analysis

Data were reduced as described in sections 2.5 and 2.7. Initial phases were calculated by molecular replacement using Phaser[45], with PDB ID *4YUM* as the search model. Prior to initial atomic refinement, R-free-flags were carried over from PDB ID *4YUM* and random displacements (σ=0.5 Å) were applied to the atomic coordinates to help remove model bias. Next, iterative cycles of model building and further refinement were performed until the models reached convergence. Individual atomic coordinates, atomic displacement parameters (B-factors), and occupancies were refined using *phenix.refine*[46,47]. Automatic identification of ordered solvent was performed during the early cycles of model refinement. Models and maps were visualized and rebuilding steps were performed using *Coot* [48]. Final structural models were visualized using PyMOL software[49], and were also used for ensemble refinement using *phenix.ensemble_refine* [50]. Input parameters for ensemble refinement (*ptls*, *tx*, *wxray_coupled_tbath_offset*) were optimized for each dataset.

Results

Optimization of Batch Crystallization

Large CypA crystals (**Figure 1.8**), on the order of hundreds of microns, or even millimeters, are readily obtained by vapor diffusion methods[26,27]; however, these crystals are too large for either microfluidic serial crystallography or MicroED. Therefore, we sought to optimize the production of microcrystals in batch format, so that they could be easily delivered to the X-ray beam for serial crystallographic measurements using microfluidic sample injectors. As a first step towards this goal, we systematically explored the phase space of CypA crystallization in the vicinity of the conditions that yield large crystals (protein concentration in the range of 80-100 mg/mL with 20-25% (w/v) PEG-3350 as a precipitant and HEPES buffer at pH 7.5). We adapted the established CypA crystallization protocol[26] to a microbatch format (rather than vapor diffusion), and tested an array of conditions by varying protein and precipitant concentrations across the two axes of a 96-well crystallization plate. The lowest concentration of protein and precipitant led to the formation of large crystals that were ideal for data collection under traditional rotation conditions. In microbatch format, conditions that resulted in large, single crystals contained substantially lower protein and precipitant concentrations relative to vapor diffusion experiments that yield similarly sized crystals. Increasing protein concentration led to the formation of a greater number of smaller crystals; however, they tended to cluster together and displayed a needle-like morphology (**Figure 1.1**). High protein concentrations also led to a large variation in crystal size, which we sought to avoid since crystal monodispersity is desirable for serial crystallography experiments. In addition to modulating protein concentration, precipitant concentration was also varied. Increasing the concentration of the precipitant led to increased crystal density while maintaining better monodispersity. At the highest precipitant concentrations we tested, the protein tended to aggregate, rather than crystallizing. Given these characteristics, we found that we could consistently get dense crystal slurries when the final PEG-3350 concentration was near 20% (**Figure 1.1**). Increasing protein concentration beyond

35 mg/mL did not lead to appreciable increases in crystal density, so we chose a final protein concentration of 35 mg/mL and a final PEG-3350 concentration of 20% for further optimization.

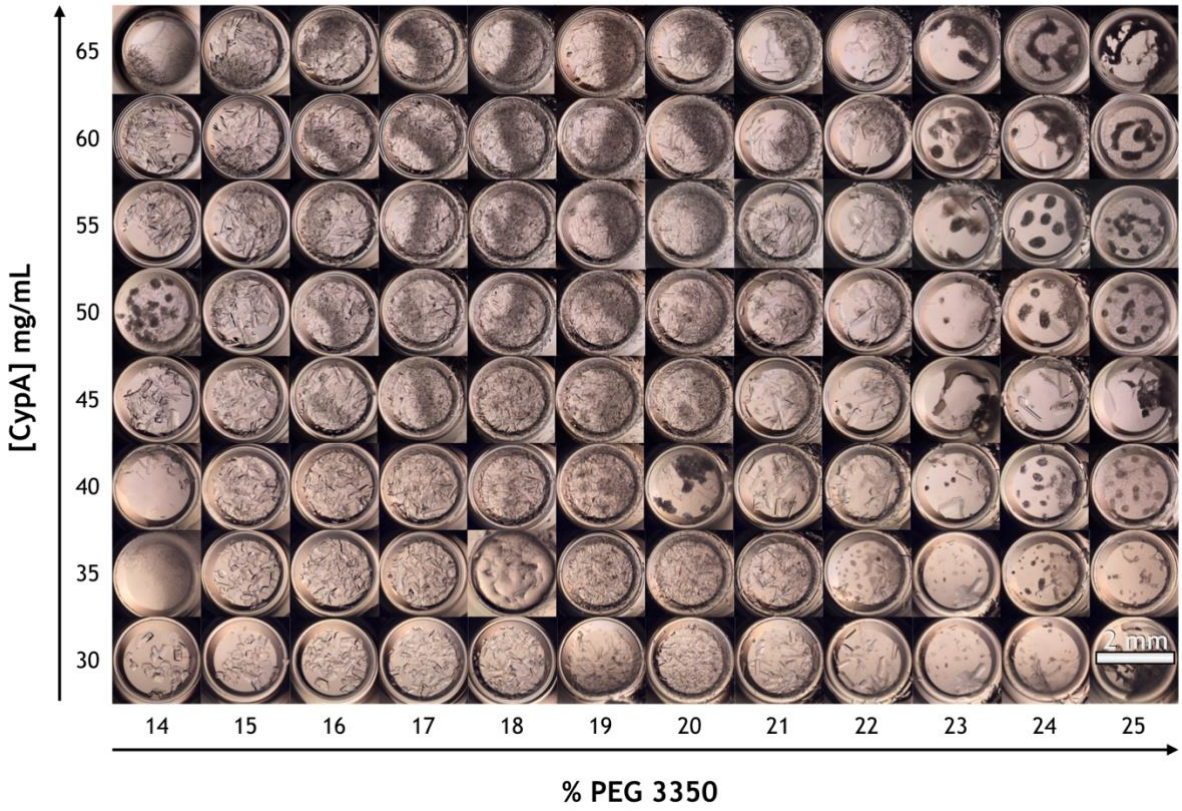

Figure 1.1. An array of images that illustrates the crystallization phase space of CypA. Concentrated solutions of CypA were mixed in a 1:1 volume ratio with solutions of varying concentrations of PEG-3350. Labels on the axes indicate final concentrations after mixing. CypA crystallizes readily in PEG-3350 solutions, however, the crystal size and morphology vary dramatically as a function of protein and PEG concentration. Specifically, at low [CypA] and low [PEG-3350] (bottom left corner), the crystals that form are few and large, while at high [PEG-3350] (right side), CypA aggregates and no crystals form. In the middle of the phase space, dense solutions of small crystals form.

After identifying ideal protein and precipitant concentrations using the microbatch method described in section 2.2, our next goal was to scale up the microbatch procedure to produce crystal slurries at the milliliter scale. We developed a batch crystallization protocol in which 0.9 mL of CypA is stirred using a magnetic stir bar inside a glass vial, and 0.6 mL of PEG-3350 solution (50% wt./vol.) is added dropwise to produce a solution with final protein and

precipitant concentrations of 36 mg/mL and 20% (wt./vol.), respectively. Initially, we used a rotating mixer to mix the slurry by inversion, but adding a stir bar yielded better results. Our batch stirring protocol also improved monodispersity, decreased crystal size, and increased crystal density (crystals / mL) relative to the microbatch method. Increasing the final protein concentration above 36 mg / mL, and increasing the final PEG-3350 concentration above 20% did not further improve monodispersity, size, or crystal density; however, we discovered that modulating the stirring speed of the crystallization solution allowed us to control the formation of differently sized crystals (**Figure 1.2**). Within the range of stir-rates we tested (200-800rpm), we observed an increase in crystal density as stir rate increased, which was coupled to a decrease in the average crystal size. Crystals 50 x 50 x 50 µm or larger developed at lower stir rates, while crystals tended towards 15 x 15 x 15 µm at higher stir rates. In addition to modulating the density of the slurry, the stirring rate also affected monodispersity[51]. At 200 RPM, we observed greater variation in the size of the crystals, while at higher stir rates the crystals were more monodisperse, but tended to clump together.

Next, optimized microcrystals were tested for their diffraction quality using a synchrotron beamline (SSRL 12-2) and exhibited Bragg peaks were visible beyond 2.0 Å resolution (**Figure 1.9**). the indexed unit cell dimensions matched those of previously published CypA structures[27], indicating that the batch crystallization method does not have an adverse effect on the quality of the crystal lattice. The crystals used for this test were 50 x 50 x 50 µm, while typical CypA crystals used for single-crystal X-ray crystallographic structure determination are over 100 µm in each of their three dimensions.

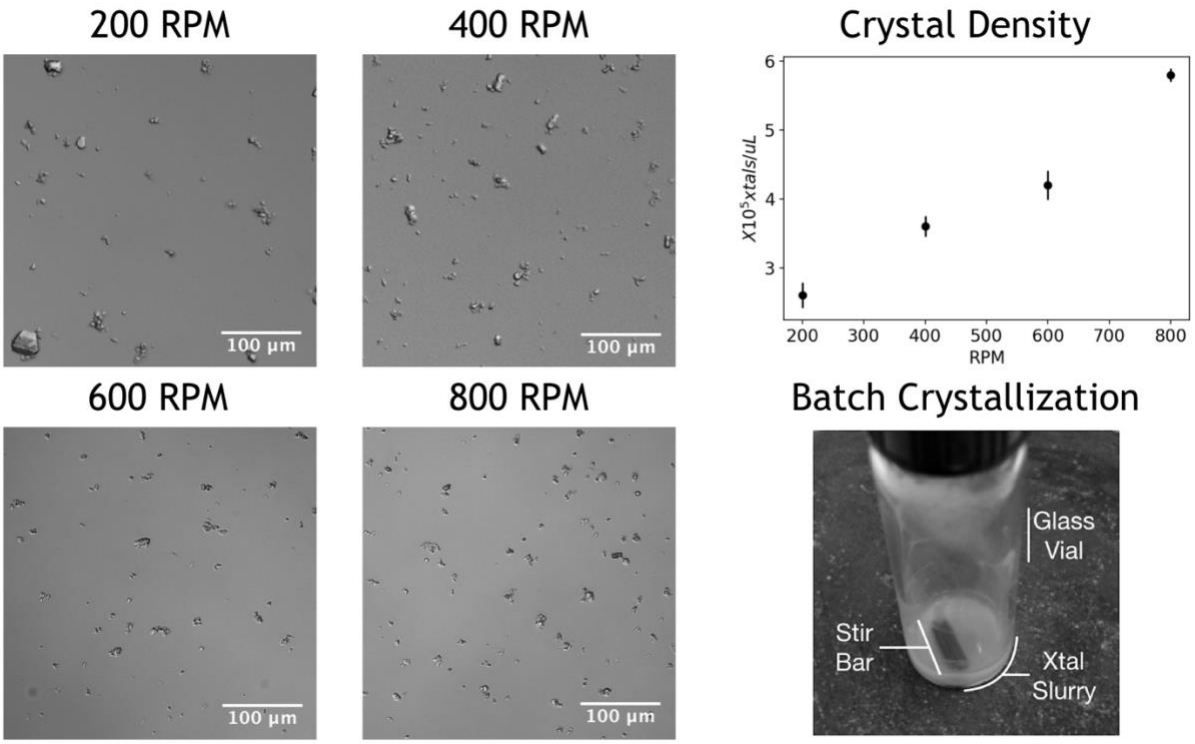

Figure 1.2. Images of microcrystals formed in batch with constant stirring. As stir rate increased, the average size of the crystals decreased, and the density of the slurry increased. This was confirmed by assessing crystal density using a hemocytometer.

Table 1.1. Sample injection parameters for serial XFEL experiments.

	MESH Injector	LCP	Cellulose
Sample Flow Rate	0.3 µL / min	.088 µL / min	0.625 µL / min
Capillary Diameter	250 µm	50 µm	75 µm
Linear Jet Velocity	Not determined	0.75 mm / s	9.43 mm / s
Electric Field	~3000 V/cm	0 V/cm	0 V/cm
Pressure	<20 PSI	~8500 PSI	~3200 PSI
Delivery Matrix	As crystallized	LCP-like lipid mixture	20% Hydroxyethyl cellulose

Serial X-ray Crystallography - Microfluidic Sample Delivery and Data Collection

We used CypA microcrystal slurries to perform several different types of serial crystallography experiments, all conducted in ambient atmosphere, in order to assess the performance of different methods of microfluidic sample delivery, and to determine whether the conditions created in the injector alters the outcome of the structural measurements. Specifically, we tested two different types of microfluidic sample injectors, one which utilizes the principle of electrospinning to form a microfluidic stream[19], and another that performs high-pressure microextrusion of crystals embedded in a viscous material[20]. Because the microextrusion method requires that samples are extremely viscous, we tested three different viscogens as additives to our CypA samples. Thus, a total of four unique experimental conditions were explored.

The first injector system we implemented, in an experiment conducted at the MFX endstation of the XFEL at the Linac Coherent Light Source (LCLS), is referred to as the microfluidic electrokinetic sample holder (MESH). This device relies upon the principle of electrospinning to break the surface tension of the crystal slurry and drive it into a microjet as it exits the tip of a capillary[19]. Within the MESH system, gentle pressure (less than 20 PSI) from a syringe pump drives crystals through a capillary (250 µm i.d.) toward the X-ray interaction point, and application of approximately 3000 V of electrostatic potential (**Table 1.1**) across the sample stretches the liquid into a thin jet as it emerges from the capillary tip. Although our crystals were much smaller than 250 µm, they had a tendency to cluster together (**Figure 1.2**), so we used a capillary with a relatively large internal diameter to avoid clogging. A full description of the optimal operating parameters for our experiment using the MESH is provided in **Table 1.1**. During this experiment, the position and physical dimensions of the Taylor cone and microfluidic jet formed by electrospinning (**Figure 1.10**) had a tendency to fluctuate. However, by positioning the injector so the X-ray beam was at the approximate position where the liquid within the Taylor cone accelerated and became the jet, we were able to collect data at a hit rate of approximately

19% and an indexing rate of 63% (**Table 1.2**). Moving the injector so that the beam was pointed at the jet itself resulted in an unacceptably low hit rate (<1%), and moving the injector so that the beam was positioned at a more stable, but thicker, region of the Taylor cone resulted in extremely high background scattering due to the excessive volume of solvent in the beam path.

The second type of microfluidic sample injector we implemented for our experiments was a viscous extrusion system. Several variations on this device have been created, all of which operate using high pressure to extrude a viscous crystal slurry through a capillary, which is stabilized by a sheath gas to form a relatively slow-moving column of material as it exits the capillary at the X-ray interaction point[20]. Within the injector, crystals were exposed to pressures as high as 8500 PSI (**Table 1.1**), with no electrostatic potential. Given the high operating pressures of these injectors, we experienced no issues with clogging, so we used a 50 µm capillary to minimize sample consumption. Because these injector systems require that the sample be much more viscous than our CypA crystal slurries, their use required that we add viscogens, or "carrier media," to our samples after crystallization but prior to injection. We experimented with three different types of carrier media: lipidic cubic phase (LCP) formed from monoolein, hydroxyethyl cellulose, and polyethylene oxide (PEO). Carrier media solutions were prepared independently and were mixed with crystal slurries to embed the microcrystals in the viscous material immediately prior to loading samples into the injector reservoirs. For LCP and PEO, the carrier media and crystal slurries we prepared in separate glass syringes, and mixed through a coupling device, while samples containing cellulose were prepared by mixing crystal slurry and carrier medium on a glass surface using a spatula[52]. Visual inspections, using a microscope equipped with a cross-polarizer, confirmed that while it seems harsh, the process of mixing CypA microcrystals into viscous material does not visibly damage them. Crystal slurries prepared with LCP and PEO carrier media were delivered to the XFEL interaction point of the MFX endstation at the Linac Coherent Light Source (LCLS) using an injector device developed by[20]. Crystal slurries prepared with hydroxyethyl cellulose were delivered to the XFEL

interaction point of the SPring-8 Angstrom Compact Linear Accelerator (SACLA), using an injector setup similar that used in studies Photosystem II[14,53,54]. We observed that samples prepared with both LCP and hydroxyethyl cellulose formed microfluidic jets that were highly stable and maintained consistent physical dimensions for long periods of time, allowing efficient data collection. We obtained an average hit rate of 20% and an indexing rate of 30% for our experiment with LCP as the carrier medium, and we obtained an average hit rate of 20% and an indexing rate of 75% for our experiment with cellulose as the carrier medium (**Table 1.2**). On the other hand, and in contrast to reports by[55], we were unable to obtain useful data when PEO was used as the carrier medium. Samples prepared using PEO did not form stable jets, but instead formed droplets at the tip of the injector nozzle, which grew to a critical mass and then slowly dripped toward the X-ray interaction point and became unstable in the sheath gas. This problem persisted, despite considerable effort to optimize the sample and sheath gas flow parameters, and attempting to use both helium and nitrogen as the sheath gas.

Table 1.2. Crystallographic statistics for data collection. Statistics for the highest-resolution shell are shown in parentheses.

	XFEL MESH	XFEL LCP	XFEL Cellulose
X-ray Source	LCLS-MFX	LCLS-MFX	SACLA-BL2
Photon Energy (keV)	9.4	9.5	10
X-ray pulse duration (fs)	~50	~45	~10
photons/pulse	~$1*10^{12}$	~$2.5*10^{12}$	~$1*10^{11}$
Repetition Rate (Hz)	10	10	30
Hit Rate (%)	19	20	20
Indexing Rate (%)	63	30	75
No. of Images	18358	11821	23947
Resolution Range	20.0-1.62 (1.65-1.62)	20.0-1.65 (1.68-1.65)	20.0-1.56 (1.58-1.56)

	XFEL MESH	XFEL LCP	XFEL Cellulose
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell	43.32±0.11 Å 52.94±0.09 Å 89.87±0.21 Å 90° 90° 90°	43.10±0.18 Å 52.65±0.13 Å 89.29±0.26 Å 90° 90° 90°	43.00±0.26 Å 52.60±0.23 Å 89.20±0.37 Å 90° 90° 90°
Total Reflections	5928801 (37371)	2477563 (11714)	7874934 (59803)
Multiplicity	224.08 (28.81)	98.90 (9.63)	261 (40.5)
Completeness (%)	99.95 (100)	99.89 (99.92)	100 (100)
Mean I/sigma(I)	5.312 (0.747)	4.062 (1.076)	7.91 (1.46)
Wilson B-factor	20.21	19.84	24.82
R-split	8.3 (28.4)	14.9 (54.3)	7.94 (66.0)
CC-int	99.3 (86.5)	94.5 (56.7)	99.3 (58.6)
Software	cctbx+cxi.merge	cctbx+cxi.merge	CrystFEL

In addition to assessing the data quality under different sample delivery conditions, we also wanted to determine whether the different types of sample delivery methods bias the orientation of the crystals as they are delivered to the X-ray beam (**Figure 1.11**). Because serial crystallography methods assume that crystals are delivered to the beam in random orientations in order to sample all of reciprocal space, the extent to which this is not true limits the efficiency of the experiment. We found that for the MESH data set, crystals do not appear to have an orientation bias as they are delivered to the X-ray beam, while the data set collected using LCP as a carrier medium for a viscous extrusion injector showed significant orientation bias.

Serial X-ray Crystallography - Data Quality and Atomic Structure are Robust Across Sample Delivery Strategies

Following data collection, the individual data sets were processed, and the reduced data were compared, revealing that the high-quality diffraction typical of CypA crystals is consistent across the different sample delivery methods that we implemented in our serial crystallography experiments. Raw diffraction images collected at LCLS (MESH and LCP conditions) were indexed and integrated using *dials.stills_process*[32], and the individual measurements were merged (with post-refinement) using *cxi.merge*[34]. Raw diffraction images collected at SACLA (cellulose condition) were indexed, integrated, and merged using CrystFEL, following hitfinding with Cheetah. We used different software to process data obtained at different XFEL light sources because we observed that optimal software performance generally depends on idiosyncratic features of the experimental endstations, such as detector behavior and spectral characteristics of the X-ray pulses. The data sets were comprised of 18358, 11821, and 23947 indexed diffraction images each for the MESH, LCP, and cellulose samples respectively. The large number of indexed patterns used to construct each data set resulted in very high completeness and multiplicity, and CypA crystals diffracted to high resolution under all three delivery strategies (**Table 1.2**). The diffraction resolution of the data sets reported here fall within the range of resolutions reported for CypA structures solved using large crystals and rotation geometry, and the modest differences in maximum resolution (1.65-1.56 Å) between the data sets are likely due to differences in the number of indexed patterns contributing to each data set, rather than to significant differences in the quality of diffraction under the three distinct sample delivery conditions. Statistical metrics reflecting the measurement precision and strength of the diffraction signal were favorable for all data sets, indicating that none of the delivery methods compromised the integrity of the crystal lattice (**Table 1.2**).

Using the reduced data sets, we performed molecular replacement to calculate initial phases, followed by iterative cycles of model building and atomic refinement to determine the

structure of CypA under each of the different sample delivery conditions. After molecular replacement and before the initial cycle of manual model building, we applied random perturbations (σ = 0.5 Å) to the atomic coordinates and refined them against the X-ray data in order to eliminate any effect of model bias that might arise from using the same molecular replacement search model for the three independent structures. We performed iterative rounds of model building and atomic refinement until the procedure reached convergence, and found that the models we obtained from each of the three experiments were of comparable statistical quality in terms of their fit to the experimental data and overall geometry (**Table 1.3**). We discovered that the method of sample delivery in each of the three serial crystallography experiments has minimal impact on the average structure of CypA; however, the three individual structures are not identical.

Table 1.3. Statistics for X-ray model refinement. Statistics for the highest-resolution shell are shown in parentheses.

	XFEL MESH	XFEL LCP	XFEL Cellulose
Resolution Range	19.91-1.62 (1.72-1.62)	17.43-1.65 (1.75-1.65)	19.9-1.56 (1.64-1.56)
Unique Reflections	26445 (4324)	25034 (4076)	29528 (4159)
Reflections used in Refinement	25,613 (4189)	24247 (3950)	28,598 (4027)
Reflections used for R-free	832 (135)	787 (126)	929 (132)
R-work	0.1362 (0.2398)	0.1434 (0.2316)	0.1348 (0.2376)
R-free	0.1569 (0.2455)	0.1671 (0.2367)	0.1509 (0.2804)
Number of non-hydrogen atoms	1522	1558	1551
macromolecules	1383	1399	1401
Protein residues	163	163	163
RMS (bonds)	0.005	0.004	0.015
RMS (angles)	0.844	0.656	1.298
Ramachandran favored (%)	96.89	96.89	96.89

	XFEL MESH	XFEL LCP	XFEL Cellulose
Ramachandran allowed (%)	3.11	3.11	3.11
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	0.68	2.68	0.00
Clashscore	3.24	1.43	1.78
Average B-factor	26.29	26.78	29.62
macromolecules	24.78	25.29	28.02
PDB ID	6U5C	6U5D	6U5E
Ensemble Refinement			
R-work	0.1241	0.1304	0.1296
R-free	0.1477	0.1517	0.1512

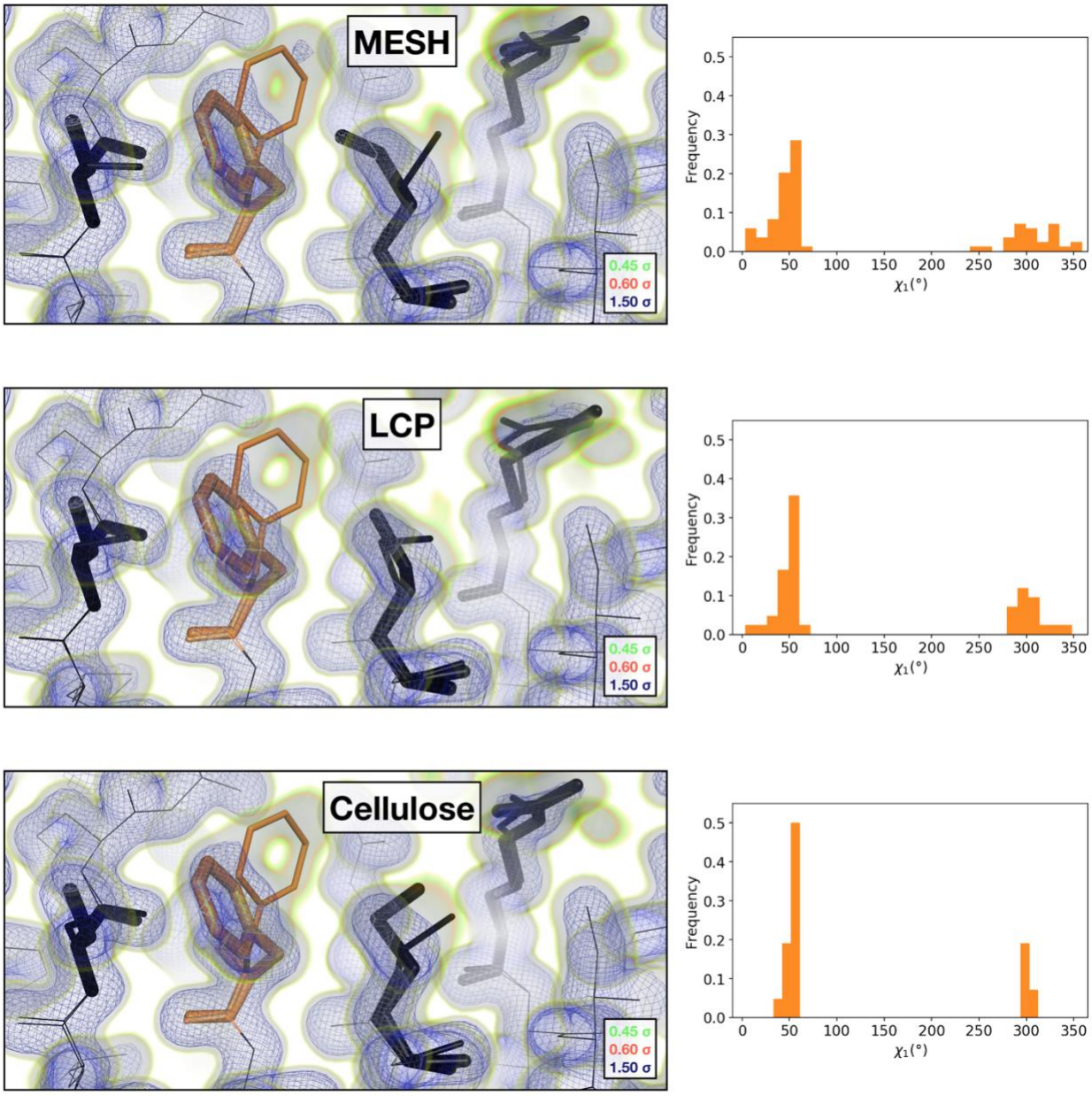

Figure 1.3. Comparison of the 2mFoFc map and the refined multi-conformer model produced from each serial XFEL experiment. Maps were visualized at multiple contour levels to show evidence for alternative conformations. Following multi-conformer refinement, ensembles were generated from each model using *phenix.ensemble_refine*. In the right panel, a histogram of the Chi1 angles for residue 113 is plotted for the ensemble. Multi-conformer models plus maps, and the distribution of Chi1 angles across the ensemble models are similar for all three XFEL data sets.

Pairwise alignment of the three structures and comparison of atomic coordinates revealed that for each pair of structures, the root-mean-squared deviation (RMSD) of atomic positions is less than 0.1 Å (**Table 1.4**). Conformational heterogeneity of a key network of residues in CypA that extends from the core of the protein to the active site has been studied previously using ambient-temperature crystallographic experiments. Rotameric exchange of residues in this network, which includes R55 (the catalytic residue), M61, S99, and F113, is required for enzymatic turnover[26,56]. Notably, all three of our serial datasets revealed evidence for multiple conformations of these residues (**Figure 1.3**). Small differences between the structures existed as differences in rotamers (or mixtures of rotamers) for side chains with generally weak electron density, such as Met61.

Table 1.4. **All-atom RMSD values for comparison of the three serial crystallography structures.**

Pair	RMSD
MESH/LCP	0.048 Å
MESH/Cellulose	0.063 Å
LCP/Cellulose	0.069 Å

Furthermore, CypA's structure did not appear to be significantly perturbed by either the electric field within the MESH jet, or by the high pressures within the LCP injector used for the LCP and Cellulose datasets (**Figure 1.3**). Local features within the models matched the maps well, with only subtle differences noticeable in the maps. Model statistics were similar across all three datasets, only the average B-factor differed between appreciably between datasets (**Table 1.3**). Comparing the normalized atomic B-factors of atoms within each structure (**Figure 1.12**) revealed that the increase in the average was not due to any localized change in conformational heterogeneity, but instead resulted from a global increase in the refined B-factors. These global differences in atomic B-factors across structures could be due to varied perturbation of the

crystal lattice (but not the molecular structure) that results from exposure to different sample delivery conditions, or due to small differences in the data processing parameters. When the structure was expanded to an ensemble, the resulting multiconformer models settled into nearly equivalent minima, confirming the similarity of the three datasets (**Figures S5 & S6)**.

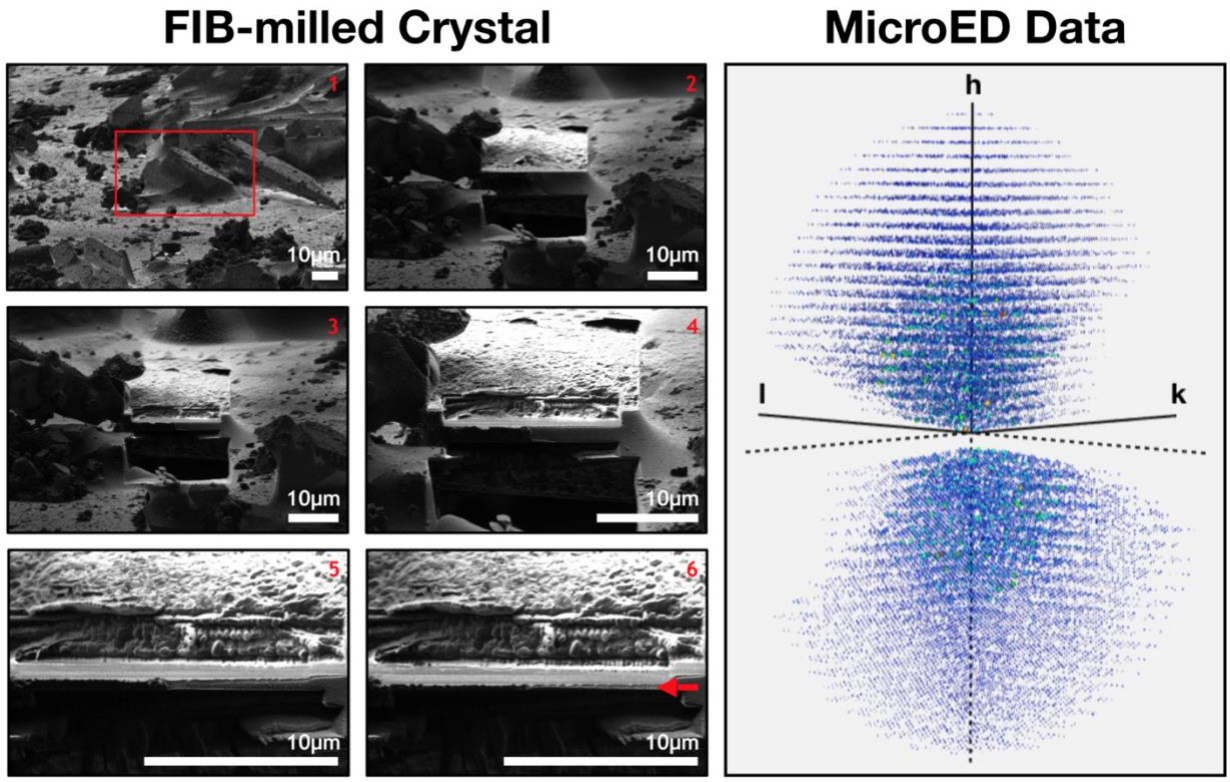

Figure 1.4. MicroED Data collection. CypA crystals were deposited on a copper grid with an amorphous carbon support material and frozen in vitreous ice (left). A single crystal (highlighted in a red box in image 1) is shown at various stages (1-6) of the FIB milling process. The edge of the final crystalline lamella is denoted with a red arrow in image 6. Also shown is the intensity weighted reciprocal lattice (right) representing the MicroED data that was collected from the single crystal shown in the left panel.

MicroED - Grid Preparation and Data Collection

In order to obtain CypA microcrystals on copper grids, suitable for MicroED data collection, we tested several sample preparation strategies. The ideal crystal thickness for MicroED samples is approximately 300-500 nm[25], which is smaller than any crystal that is

visible using light microscopy. First, we prepared grids using a CypA microcrystal slurry containing visible crystals on the order of 10 um, with the hope that this sample would also contain much smaller crystal fragments that would be acceptable for data collection. We examined this sample in the microscope and observed only large (several microns or larger) microcrystals on the grid. As a next step, we attempted to reduce the size of the CypA microcrystals using several physical agitation methods, including vortexing and crushing the crystals using either a pipette tip or glass beads. Samples exposed to physical agitation were used for grid preparation and were examined under the microscope, again revealing an absence of suitably sized crystals for data collection. Attempts to improve the grid preparation by changing the glow discharge and blotting methods also did not result in suitable samples. We hypothesize that difficulties in preparing grids with sub-micron sized CypA crystals result from a combination of the crystals' surface properties and the strong lateral forces that are introduced by the blotting process, which could pull small crystals off of the grid. We note that we used only grids with amorphous holey carbon supports, and did not attempt to prepare samples using grids with more exotic support materials, such as gold or graphene oxide.

Because we were unsuccessful in preparing samples for MicroED using traditional methods of applying crystals to grids, we turned to a method that utilizes a focused ion beam (FIB) to mill larger crystals down to an appropriate thickness. We observed that crystals larger than several microns were able to stick to the holey carbon grids, we prepared a frozen grid with crystals that were approximately 5 x 5 x 5 µm to 20 x 20 x 20 um. Prior to MicroED data collection in the transmission electron microscope (TEM), the frozen samples were loaded into a scanning electron microscope (SEM) equipped with a gallium ion FIB. A single crystal, approximately 20 x 20 x 20 µm was identified in the SEM and was subsequently FIB-milled to form a crystalline lamella that was approximately 200 nm thick (**Figure 1.4**). The grid containing the lamella was then transferred to the TEM for MicroED data collection. Three separate data sets were collected from unique regions of the lamella, two of which were merged to produce

the final reduced data set that was used for structure determination. Inclusion of the third data set degraded the quality of the merged data. Also, because the rotation range of the microscope stage is restricted, and the crystal orientation was the same for all three data sets due to the fact that they were collected from a single lamella, the final merged data set suffered from a missing wedge of reciprocal space (**Figure 1.4 and Figure 1.14**) and had an overall completeness of 86%. Bragg peaks were measured out to 1.9 Å, but integrating data beyond the 2.5A cutoff that we implemented resulted in poorer data quality. Additional information about the quality of the merged MicroED data is provided in **Table 1.5**.

Table 1.5. Crystallographic statistics for MicroED data. Statistics for the highest-resolution shell are shown in parentheses.

	MicroED
No. of Images	135
Resolution Range	30.5-2.50 (2.57-2.50)
Space Group	P $2_1 2_1 2_1$
Unit Cell	42.40 Å 53.40 Å 87.76 Å 90° 90° 90°
Total Reflections	22370 (1668)
Multiplicity	1.95 (1.94)
Completeness (%)	86 (87.3)
Mean I/sigma(I)	3.23 (1.01)
Wilson B-factor	35.53
R-meas	24.9 (87.7)
CC 1/2	95.2 (44.8)
Unique Reflections	6236 (608)
Reflections used in Refinement	6236 (608)

	MicroED
Reflections used for R-free	213 (22)
R-work	0.1854
R-free	0.2237
Number of non-hydrogen atoms	1280
macromolecules	1248
Protein residues	163
RMS (bonds)	0.011
RMS (angles)	0.754
Ramachandran favored (%)	96.27
Ramachandran allowed (%)	3.73
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.76
Clashscore	0.41
Average B-factor	30.60
macromolecules	30.97
PDB ID	6U5G
Ensemble Refinement	
R-work	0.2351
R-free	0.2587

CypA Structure Determination from MicroED Data

After merging the integrated MicroED data to obtain a high-quality data set, we implemented the exact same structure determination procedure as was used for analysis of the serial X-ray diffraction data sets. Specifically, we performed molecular replacement followed by application of random coordinate perturbations, and then iterative model building and refinement until the Rwork and Rfree values converged and no additional improvements to the model could

be made. Analysis of the CypA crystal structure determined by MicroED revealed two notable features.

First, during the indexing stage of the data reduction procedure, we observed that the unit cell had unusual dimensions (**Figure 1.5**). Specifically, while the crystallographic a and c axes match well to those of other CypA structures determined at cryogenic temperatures, the b-axis is approximately 1% longer than the corresponding axes in typical CypA structures determined at "physiological" (>260 K) temperatures. Because of indexing challenges resulting from the systematic incompleteness of the data (14% of predicted observations missing), such as the inability to observe reciprocal lattice points along the principal k and l axes within the 60 degree wedge that we measured (**Figure 1.4**), we took several additional steps to ensure that the unit cell was indexed accurately and that the space group symmetry was correctly assigned. To address the possibility that an optical distortion in the microscope or challenges related to the flatness of the Ewald sphere[57] could lead to incorrect measurement of unit cell dimensions, we performed a structure refinement procedure that also simultaneously refined the coordinates and the lengths of the unit cell axes[58] using REFMAC5[59,60]. This procedure resulted in refined unit cell dimensions that were similar to the input (**Table 1.6**), without improvement in Rwork and Rfree, indicating that the refined unit cell does not result in a model that is more consistent with the measured intensities. To confirm that the elongation of the b-axis does not also break the crystallographic $P2_12_12_1$ space group symmetry, we reduced the raw data three separate times in space group $P2_1$. In each of these three data sets, the two-fold symmetry operation was preserved along a different crystallographic axis (i.e. $P2_111$, $P12_11$, and $P112_1$ relative to the parent $P2_12_12_1$). Refinement of the CypA structure against the data with lower symmetry produced models with worse overall quality than when the data were reduced in space group $P2_12_12_1$, confirming the validity of the space group assignment.

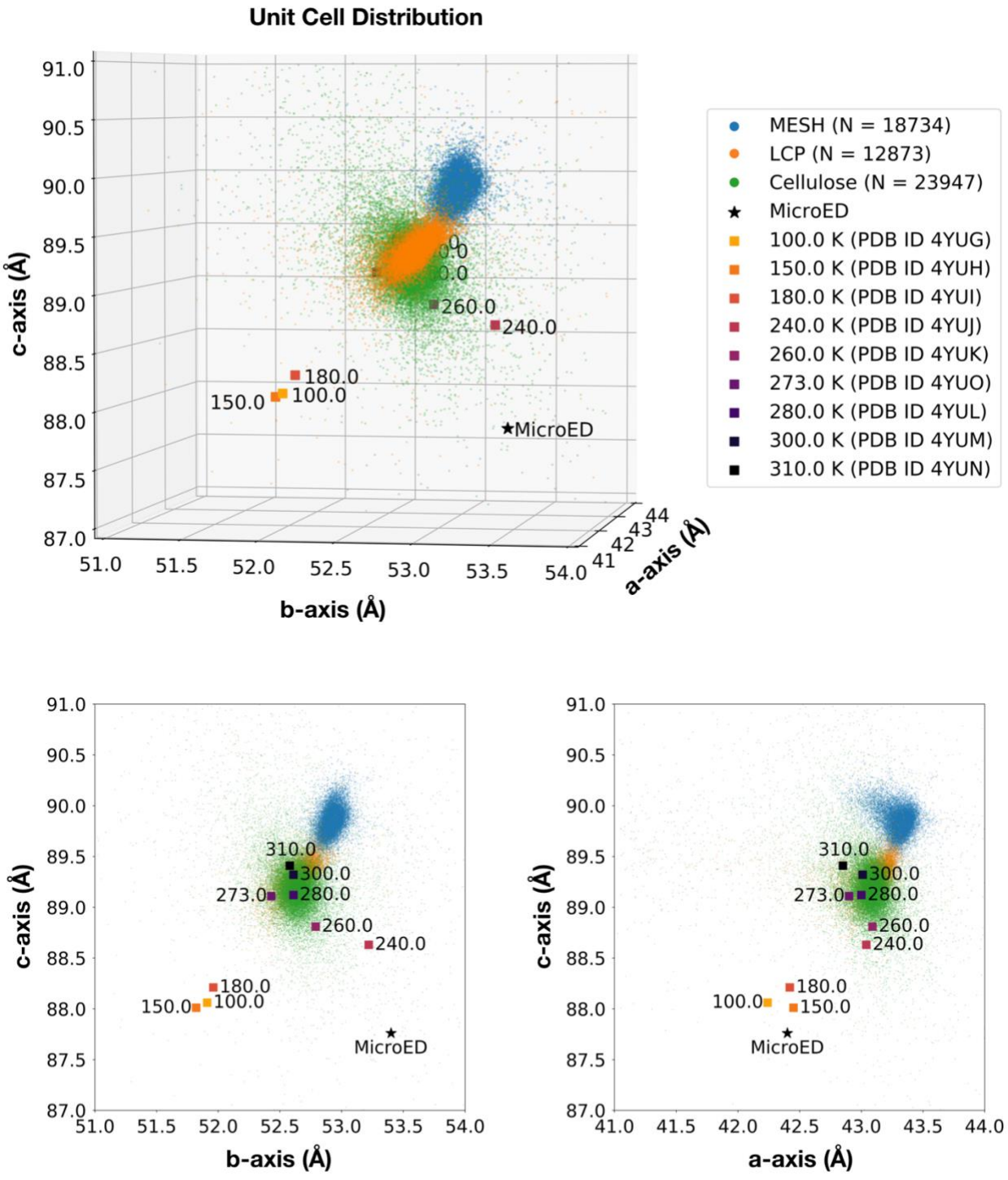

Figure 1.5. Comparison of unit cell dimensions across data collection strategies. Published structures are provided as a reference for the effect of temperature upon unit cell dimensions. The unit cells measured using serial XFEL experiments resemble data from published room temperature structures. A FIB-milled crystal used for MicroED revealed dimensions that were unique from the unit cell compression normally seen in cryogenic X-ray data.

The second notable observation that we made about the crystal structure determined by MicroED is that, while the unit cell is distorted relative to other CypA structures, the structure of the molecule within the unit cell is essentially the same as for X-ray structures that were determined using cryocooled crystals (PBD ID *3K0M*), RMSD = 0.22 Å. Structures of CypA determined from cryocooled crystals, using both X-ray and MicroED, lack key conformations that are visible in their ambient temperature counterparts. In particular, ambient temperature structures of CypA reveal alternative conformations of an important network of amino acid side chains (the catalytic residue R55, as well as M61, S99, and F113), while structures determined using cryocooled samples, including the MicroED structure presented here, reveal only a single conformation of these side chains (**Figure 1.6**). Truncating the resolution of room temperature X-ray data sets to match the MicroED resolution limit and generating electron density maps revealed that the loss of the alternate conformation in the maps calculated from MicroED data could not be attributed to the differences in resolution (**Figure 1.15**).

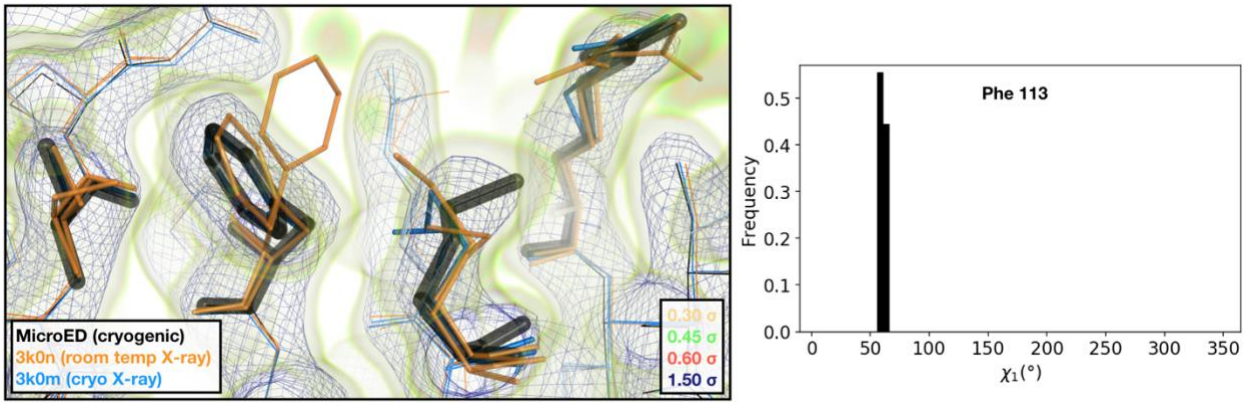

Figure 1.6. Visualization of the 2mFoFc map and the refined model measured from a FIB-milled crystal using MicroED. The conformation of residues coupled to the catalytic site resembles structures previously solved under cryogenic conditions using X-ray crystallography (PDB ID *3K0M*). For some regions of the structure, the cryogenic X-ray and MicroED structures are indistinguishable. A previously published multi-conformer model produced from data acquired at room temperature is provided for comparison (PDB ID *3K0N*). Following refinement, ensembles were generated using *phenix.ensemble_refine*. In the right panel, a histogram of the Chi1 angles for residue 113 is plotted for the ensemble. All members of the ensemble adopted the same rotameric position as previous cryogenic structures.

Effect of Experimental Conditions on Unit Cell Dimensions

Comparing the structures that we determined using different microcrystallography techniques revealed that the unit cell dimensions of the CypA crystals were noticeably affected by the conditions required for each of the different experiments (**Figure 1.5**). The three serial X-ray datasets closely resemble previous room-temperature data collected under traditional rotation conditions. Additionally, we observed that for sample delivery methods that involve embedding the crystals in a viscous carrier medium (LCP or cellulose), the unit cell tended to be slightly smaller than for experiments that do not require such additives (MESH). Average unit cell dimensions for the LCP and cellulose data sets were approximately 0.5-0.7% smaller than for the MESH data set, and the small differences between these average values were found to be statistically significant using one-way ANOVA and post-hoc Tukey tests (**Table 1.7**). We hypothesize that the observed shrinkage of the unit cell could be due to crystal dehydration. The MicroED data revealed a previously unreported unit cell, with an expanded b-axis relative to any other CypA structure that has been reported (**Figure 1.5**). The unit cell was different from both ambient temperature (PDB ID *3K0N*) and cryogenic (PDB ID *3K0M*) X-ray structures, and matches most closely to a structure determined at 240 K (PDB ID *4YUJ*). The unusual unit cell observed in the MicroED experiment is not the result of measurement error (**Table 1.6**), and we hypothesize that it could be the result of cooling the crystals in ethane, rather than nitrogen, or could be caused by the grid blotting procedure or FIB-milling, both of which are unique to MicroED. Despite the small variations in crystal packing that cause changes in unit cell parameters for structures determined using different methods, the refined coordinates of CypA molecules themselves are consistent.

Discussion

The ability to measure diffraction signals from ever smaller crystal samples has enabled a variety of new and innovative experiments in macromolecular crystallography, however there

is still a relative absence in the literature of practical guidelines for optimizing microcrystallography experiments. The work we present here attempts to address this knowledge gap, by providing a detailed description of how we optimized the growth of cyclophilin A (CypA) microcrystals and measured their diffraction using two emerging microcrystallography techniques, serial XFEL crystallography and microcrystal electron diffraction (MicroED). Our results compare and contrast serial X-ray and MicroED methodologies, and highlight some important considerations and pitfalls that might be encountered during preparation of microcrystalline samples for the respective experiments. Because we measured MicroED data from only a single crystal, we focus our comparison on technical issues related to sample preparation and ease of data collection. Consequently, this case study provides a roadmap for experimenters who are interested in performing structural measurements using crystalline samples with dimensions on the scale of nanometers to microns.

For decades, macromolecular crystallographers have strived to grow large (hundreds of microns) single crystals that can be used for crystallographic measurements using rotation X-ray methods, but new data collection methods such as serial X-ray crystallography and MicroED require the reliable formation of crystals that are much smaller, typically hundreds of nanometers to tens of microns. Precise control of crystal size over this range is challenging, and others have developed methods that employ specialized equipment for *in situ* light scattering measurements to evaluate crystal size in real time[61,62] and halt crystallization as it progresses. Instead, our work with CypA demonstrates a simple, alternative method for controlling the size of crystals during batch growth. Starting from crystallization conditions identified by microbatch screening in 96-well plates, we scaled up the crystallization volume and introduced agitation (by stirring) to control the crystal size. We observed that at higher stir rates (i.e. more agitation) crystals tend to be smaller and more concentrated in the resulting slurry. We speculate that stirring fractures the crystals when they reach a critical size, which exerts control over the

crystal dimensions and also actively introduces seeds into the slurry. For CypA, we found that batch crystallization with stirring could be used to generate relatively monodisperse crystal slurries, in milliliter volumes, with crystal sizes in the range of microns to hundreds of microns. We believe that the batch stirring protocol is likely to be useful for a variety of crystal systems beyond CypA, however it may have limited utility for crystal systems that are more susceptible to physical damage.

With the ability to create large batches of CypA crystals, we could perform serial X-ray crystallography experiments at XFEL light sources, which generally consume a large amount of sample. We utilized CypA microcrystal slurries, prepared in an identical fashion, to evaluate several commonly used, injector-based sample delivery strategies, including both electrospinning and viscous extrusion using two types of crystal carrier media. These delivery strategies exposed the crystals to extreme experimental conditions, including strong electric fields, high pressures, and unusual carrier media. Crystal structures determined using each method revealed how conditions imposed by the different sample delivery systems perturbed either the crystal lattice or the protein structure.

We first observed that the different sample delivery methods produce measurable differences in the distributions of unit cell axis lengths for CypA microcrystals. Crystals measured using the MESH device (electrospinning) tended to have longer a, b, and c axes than crystals measured using viscous carrier media. This could result from dehydration of the crystals by the viscogens (LCP or cellulose), which reduce the relative humidity of the crystallization mother liquor, or from the high pressures required to extrude the viscous carrier media through the injectors. The magnitude of the change in unit cell parameters across different sample delivery methods is similar to that which is observed for crystal cryocooling (**Figure 1.5**). Additionally, we observed that crystals delivered using the MESH (electrospinning) device tended to be oriented more randomly than crystals delivered in a viscous carrier such as LCP (**Figure 1.11**). In the case of our CypA crystals, the dihedral space group symmetry

prevents the crystals from having a dipole moment that could cause them to assume a preferred orientation in the electric field introduced by the MESH injector. On the other hand, the slightly elongated crystal morphology led to orientation bias in the high-viscosity injector system, likely due to shear forces resulting from the flow of the highly viscous liquid and the narrower inner diameter of the capillaries used in that device relative to the MESH. We expect that crystals with different properties, such as polar space group symmetry or more isotropic morphologies, would have different behaviors with respect to preferred crystal orientation in the various injector systems.

Despite differences in unit cell parameters and preferred crystal orientations, the overall quality of the reduced data sets resulting from each of the serial X-ray experiments were generally equivalent (**Table 1.3**). A notable inconsistency is that the CC1/2 and R-split values for the LCP data set are slightly worse than for the other two data sets. This could be due to the fact that the LCP data set contains the least images, or it could result from the high background created by X-ray scattering from the LCP matrix. The low resolution signal is not as strong relative to the high resolution signal for the LCP data set relative to the others, which we believe supports the latter hypothesis.

Atomic models generated from the three data sets were also strikingly similar (**Table 1.4**). Isomorphous difference maps (Fo-Fo) are a very sensitive method for detecting subtle changes in molecular structure, and we found that calculating such maps using pairs of serial X-ray data sets did produce strong (>3.0σ) peaks (**Figure 1.17**). The refined coordinates showed that these difference peaks corresponded to changes in atomic positions on the order of 0.1Å or less, which are not significant with respect to the interpretation of structure-function relationships at this resolution. We also utilized a multi-conformer ensemble refinement approach as a way to assess the level of heterogeneity (model variance) that was present in each of the data sets. Our analysis focused on a network of catalytically-important residues, which are known to be dynamic[26,56]. We observed that within the catalytic network, the refined ensembles reflect a

similar level of heterogeneity across the different structures (**Figure 1.13**), which is generally supported by the correlations in the B-factors derived from standard refinements (**Figure 1.12**). We did, however, observe that the ensemble derived from the MESH data shows enhanced heterogeneity relative to the other two data sets for a loop region including residues 69-74 (**Figure 1.16**). The conformation of this loop is stabilized by a key charged residue (R69)[63], which may be perturbed by the electric field. While it has been shown that electric fields can be used to perturb conformational dynamics in proteins[64], we expect that the effect should be minimal in our MESH experiment, because the crystals are randomly oriented relative to the applied electric field, and the field is more than two orders of magnitude less than those that are intentionally used for perturbing conformational dynamics[64].

Our results show that the choice of microfluidic sample delivery method has a minimal effect on the static crystal structure of CypA. Consequently, the choice of sample delivery method for a serial X-ray crystallography experiment should be selected based on practical considerations related to the experiment, such as the requirement for a laser perturbation or mixing in a time-resolved experiment. We note that an important aspect of microfluidic sample delivery that has not been rigorously explored is whether the effects of pressure or electric fields might have a time-dependent component as microcrystals exit the injector device and pass through the X-ray interaction region.

In addition to serial X-ray crystallography, MicroED also offers the ability to determine macromolecular structures at high resolution using very small crystals with moderate solvent content. In stark contrast to serial X-ray crystallography experiments, which require hundreds of milligrams of protein and milliliters of microcrystal slurries, MicroED lies at the other extreme, allowing crystal structure determination with as little as a single microcrystal. Additionally, MicroED experiments have a significant advantage in that they are much more accessible than experiments performed at XFELs, and require a substantially lower investment of time and resources. Using CypA crystals derived from our batch protocol, we encountered several

challenges in preparing appropriately-sized microcrystal samples on grids for measurement in the TEM. MicroED requires extremely small crystals, ideally less than 500 nm thick[25], and we encountered difficulties in getting such small CypA crystals to remain on the grids after blotting away excess solvent. This may be due to the specific surface chemistry of CypA crystals or may be a more general trend of high solvent content crystals. As a result, we turned to a recent development in sample preparation that is widening the scope of MicroED by enabling measurements from crystals that are initially tens of microns thick, by utilizing a FIB-milling process to machine large crystals into thin lamellae that are optimal for MicroED measurements[24]. The FIB-milling procedure allowed us to determine the structure of CypA by MicroED using a single crystal that was initially (before milling) similar in size to those which we used for serial X-ray experiments. We observed that the MicroED crystal structure of CypA had a slightly distorted unit cell relative to other reported CypA structures, and we hypothesize that this could be due to the damage from either blotting or FIB-milling, however more rigorous studies will be required to evaluate the specific effects of these sample prep procedures on MicroED structures. The sensitivity of the CypA crystals during preparation for MicroED could be related to their high solvent content. Our structure of CypA has the highest solvent content of any non-membrane protein MicroED crystal structure determined to date using three-dimensional crystals (**Figure 1.7**), demonstrating how improved sample preparation is expanding the technique to include more challenging crystal systems.

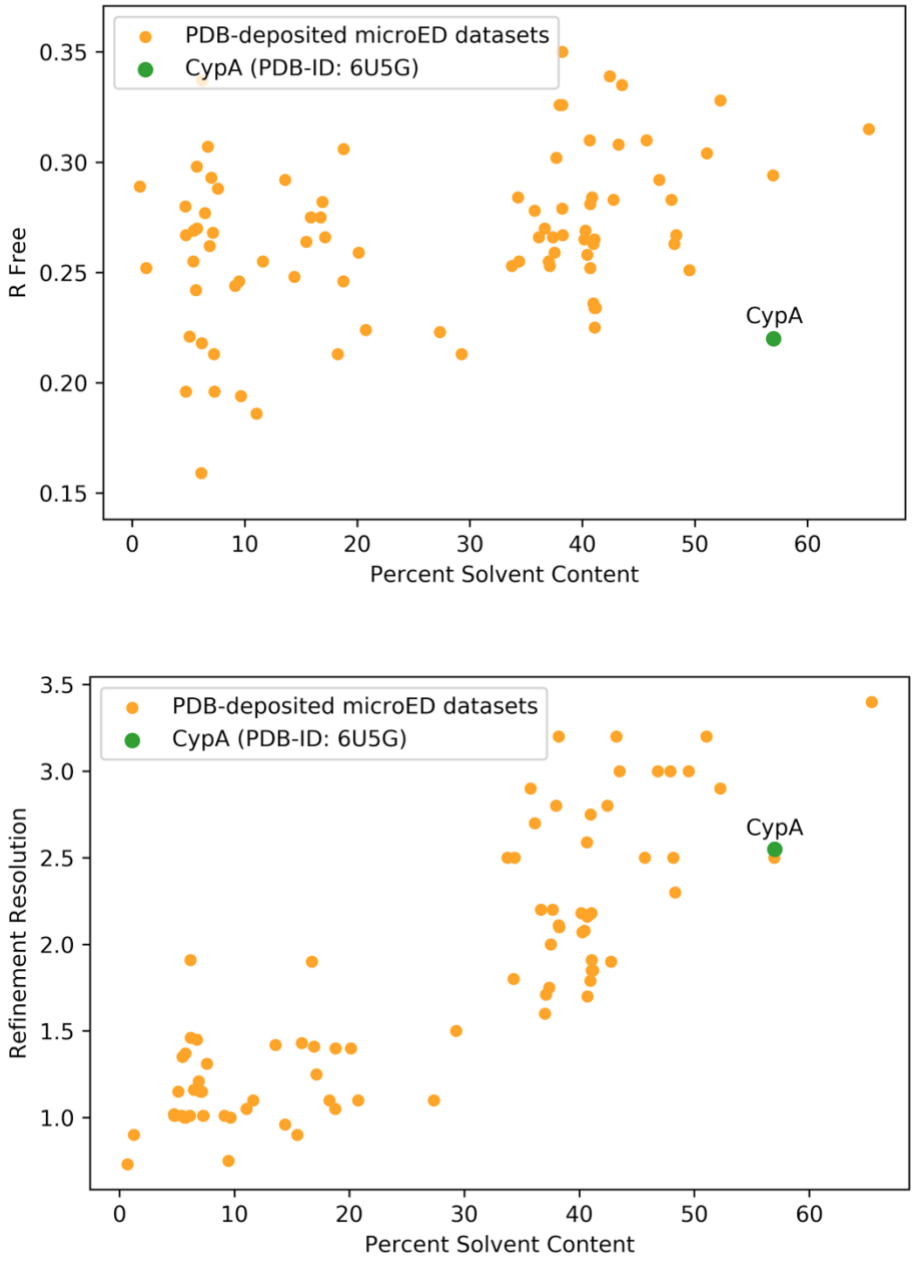

Figure 1.7. Survey of MicroED datasets deposited on the PDB. Structures determined from 3D crystals are shown as orange points, and CypA is shown as a green point. The highest solvent content point is 3J7T, which is a membrane protein.

Our evaluation of modern protein microcrystallography techniques reveals that MicroED and serial X-ray crystallography are complementary methods for structural biology[65]. The optimal experimental method for a microcrystallography experiment will depend upon various

aspects of the macromolecular system that is being studied. For determination of static, low-energy macromolecular structures, MicroED has substantial advantages over serial X-ray crystallography in terms of sample prep requirements, quantity of material required, and ease of data collection. However, while serial X-ray crystallography experiments require large amounts of sample, specialized equipment that is only available at select X-ray lightsources, and substantial optimization of sample delivery parameters, they also have their advantages. Importantly, serial X-ray measurements are performed at ambient temperature and can reveal physiological conformational ensembles of the crystallized molecules. On the other hand, we compared our MicroED structure to a cryogenic X-ray structure of CypA (*3K0M*), and observed that it was nearly identical and suffered from the same temperature-dependent reduction in conformational heterogeneity (**Figure 1.6**). Because cooling rate is related to crystal size[66], it remains to be seen whether MicroED experiments using very small crystals (hundreds of nanometers) might capture a more physiological conformational ensemble. Our data do not shed light on this question, since the crystals used in our experiments were approximately 20 x 20 x 20 µm at the time of freezing, before they were FIB-milled to an appropriate thickness. Finally, we note that for CypA, as well as other examples from the literature including lysozyme[9,10] and proteinase K[67], refinement R-factors are much higher and resolutions are generally lower for MicroED structures than for X-ray structures. In our case, some of this might be improved by collecting more complete data over multiple FIB-milled crystals. However, more generally, we expect that this discrepancy will only improve as we gain a better understanding of how electrons interact with macromolecular crystals and develop data analysis software that handles processing of MicroED data and refinement of structural models based on electron scattering more appropriately.

Acknowledgements

We thank J. Rodriguez and D. Hekstra for helpful insight. FIB milling was done in the Beckman Institute Resource Center at the California Institute of Technology. M.C.T. is supported by NSF STC-1231306, a Ruth L. Kirschstein National Research Service Award (F32 HL129989), and the UCSF Program in Breakthrough Biomedical Sciences. J.S.F. is supported by a Packard Fellowship from the David and Lucile Packard Foundation, NIH GM123159, NIH GM124149, UC Office of the President Laboratory Fees Research Program LFR-17-476732, and NSF STC-1231306. N.K.S. is supported by NIH GM117126. S.I. is supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from the Japan Agency for Medical Research and Development (AMED). R.A.W. was supported by the NSF Graduate Research Fellowship. Portions of this research were carried out at the Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory, supported by the DOE Office of Science, OBES under contract DE-AC02-76SF00515. The HERA system for experiments at MFX was developed by Bruce Doak and funded by the Max-Planck Institute for Medical Research. Portions of this research were performed at beamline 3 of SACLA with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (proposal no 2017B8055). We thank the staff at SACLA for their assistance. Data processing was performed in part at the National Energy Research Scientific Computing Center, supported by the DOE Office of Science, Contract No. DEAC02-05CH11231.

Supplementary Figures

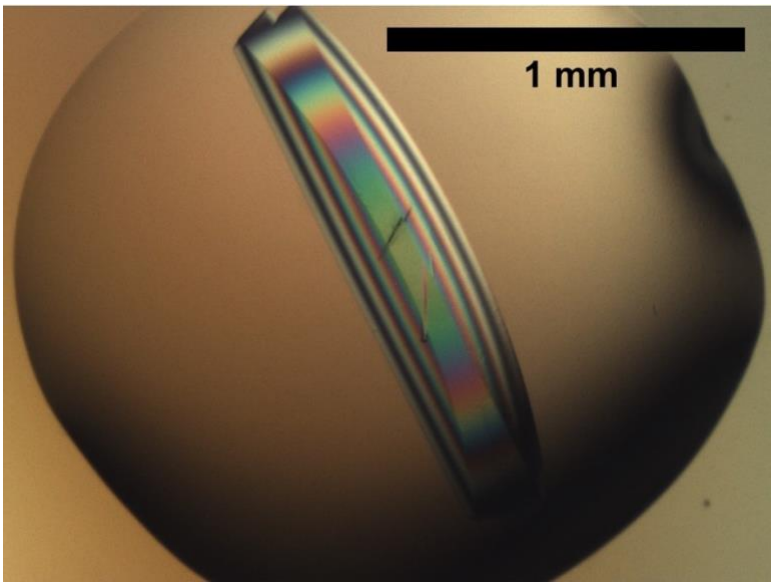

Figure 1.8. **CypA Macrocrystal.** Previously, CypA crystallization conditions were optimized to produce very large single crystals. The pictured crystal (viewed using a cross-polarizer) was over 1 mm long and was visible to the naked eye.

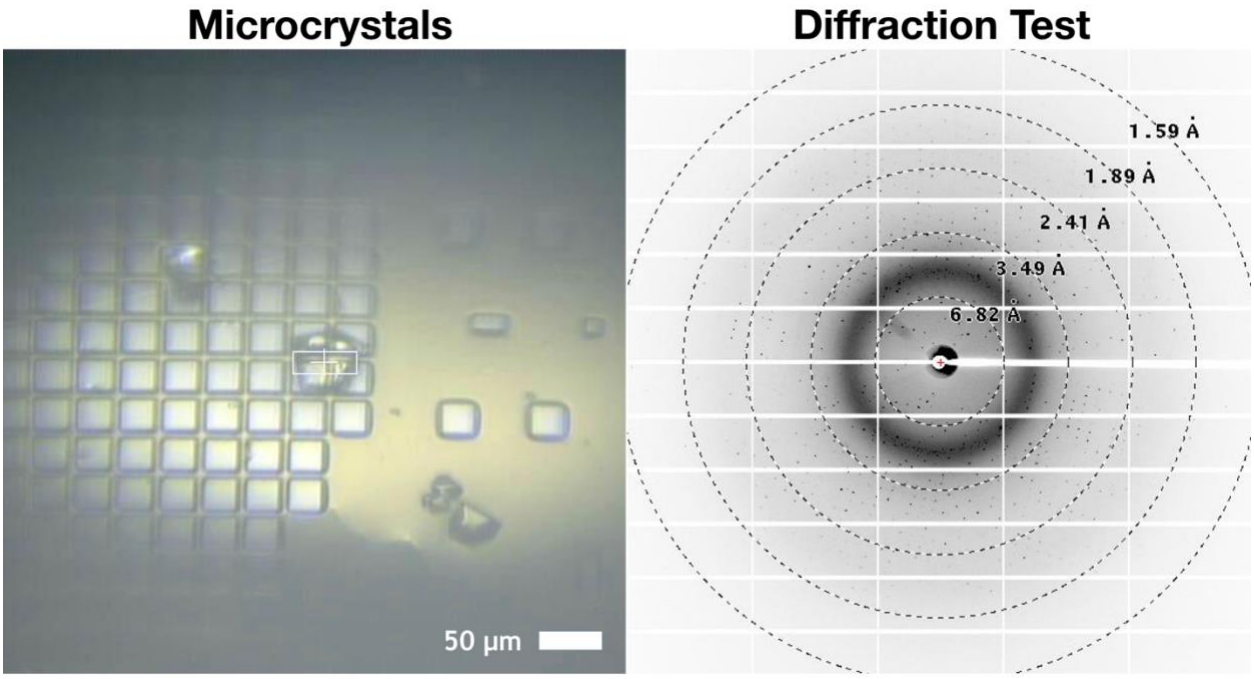

Figure 1.9. Image of microcrystals on a Mitegen micromesh grid (left) and the subsequent diffraction from one of these crystals (right). Crystals were 20-50 µm in all dimensions, and diffracted to the edge of the detector at 1.59 Å.

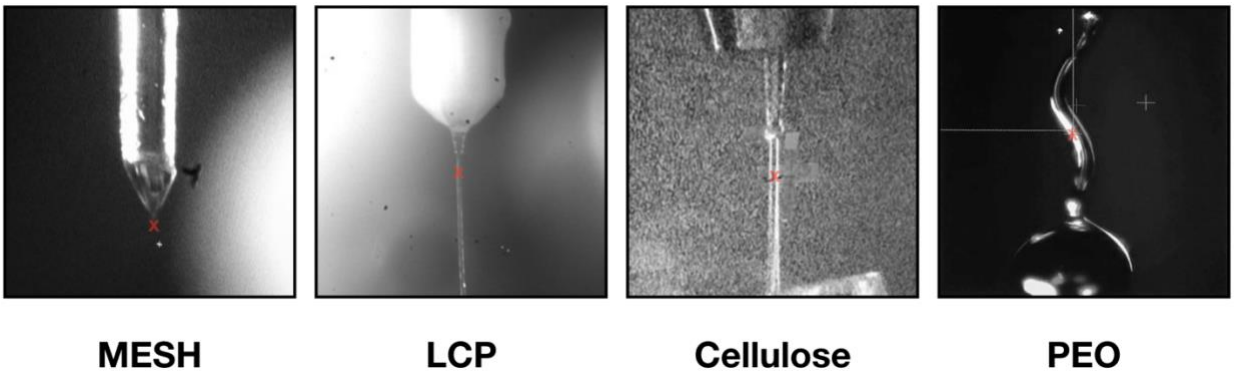

Figure 1.10. Images of jets delivering microcrystal slurries to the XFEL interaction point (red “x” in each image). Minimal viscogens were added to the crystalline slurry for the MESH injector system. When using a viscous extrusion type of injector, a variety of carrier media were tested, including LCP, Cellulose, and PEO.

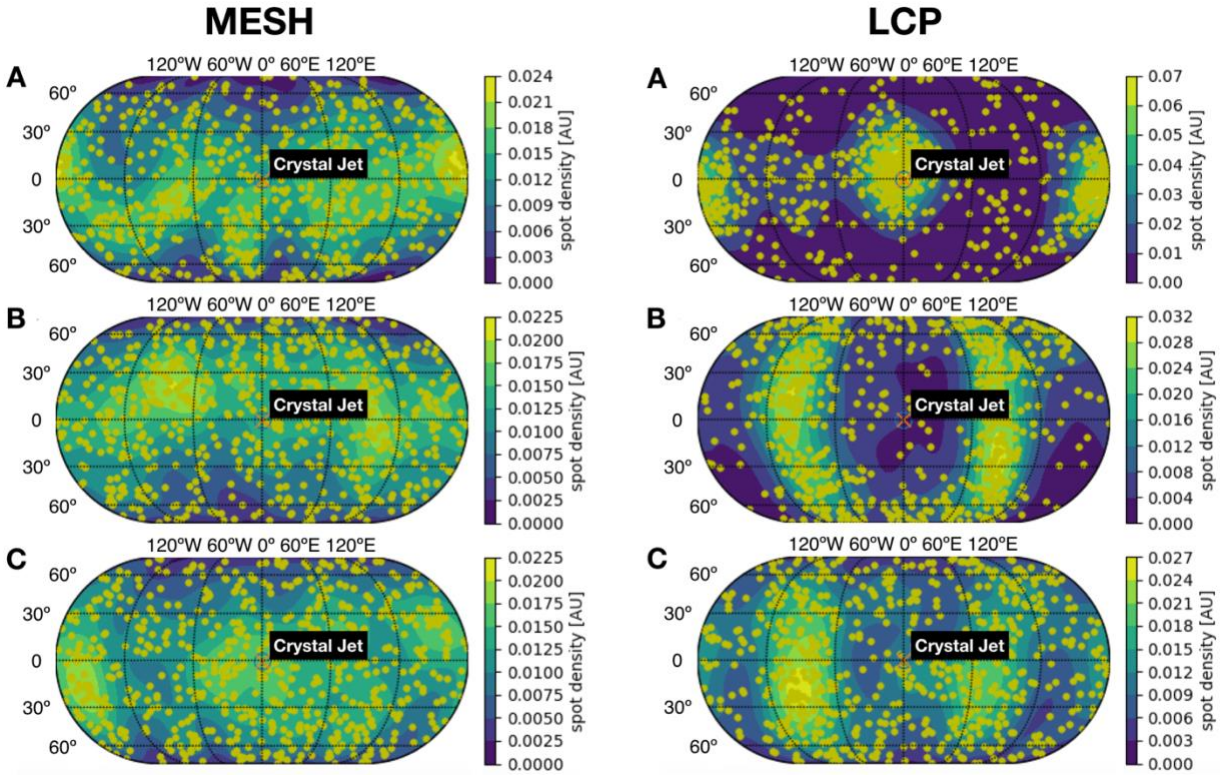

Figure 1.11. Maps of crystal axis (A, B, C) orientations from the MESH data collection (left panel) and from the LCP data collection (right panel). Axis orientations (in polar coordinates) for individual crystals are depicted as yellow spots, and the background color reflects the frequency with which a particular orientation was observed. A subset of each dataset is shown for visual clarity. The orientations appear evenly distributed for the MESH data, with no major bias introduced by the electric field created by the injection system. The viscosity of the LCP carrier media appears to have induced an orientation bias, but did not prohibit collection of a complete dataset with high redundancy.

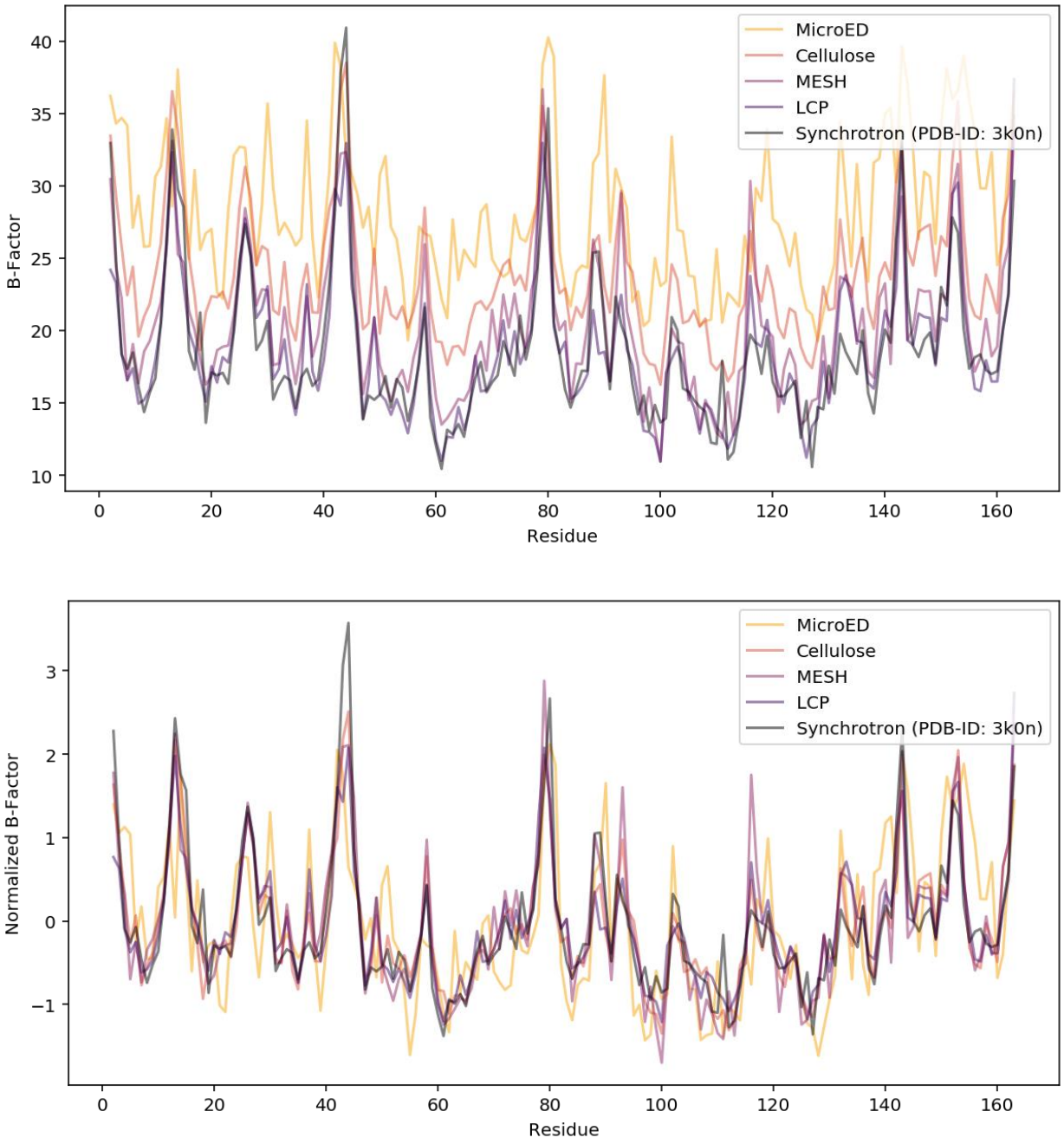

Figure 1.12. Raw (top) and normalized (bottom) B-factor per residue across data sets. A previously published structure solved at room temperature using rotational collection from a single crystal (PDB ID *3K0N*) is provided for reference. Most variation is systematic, and thus is removed by normalization.

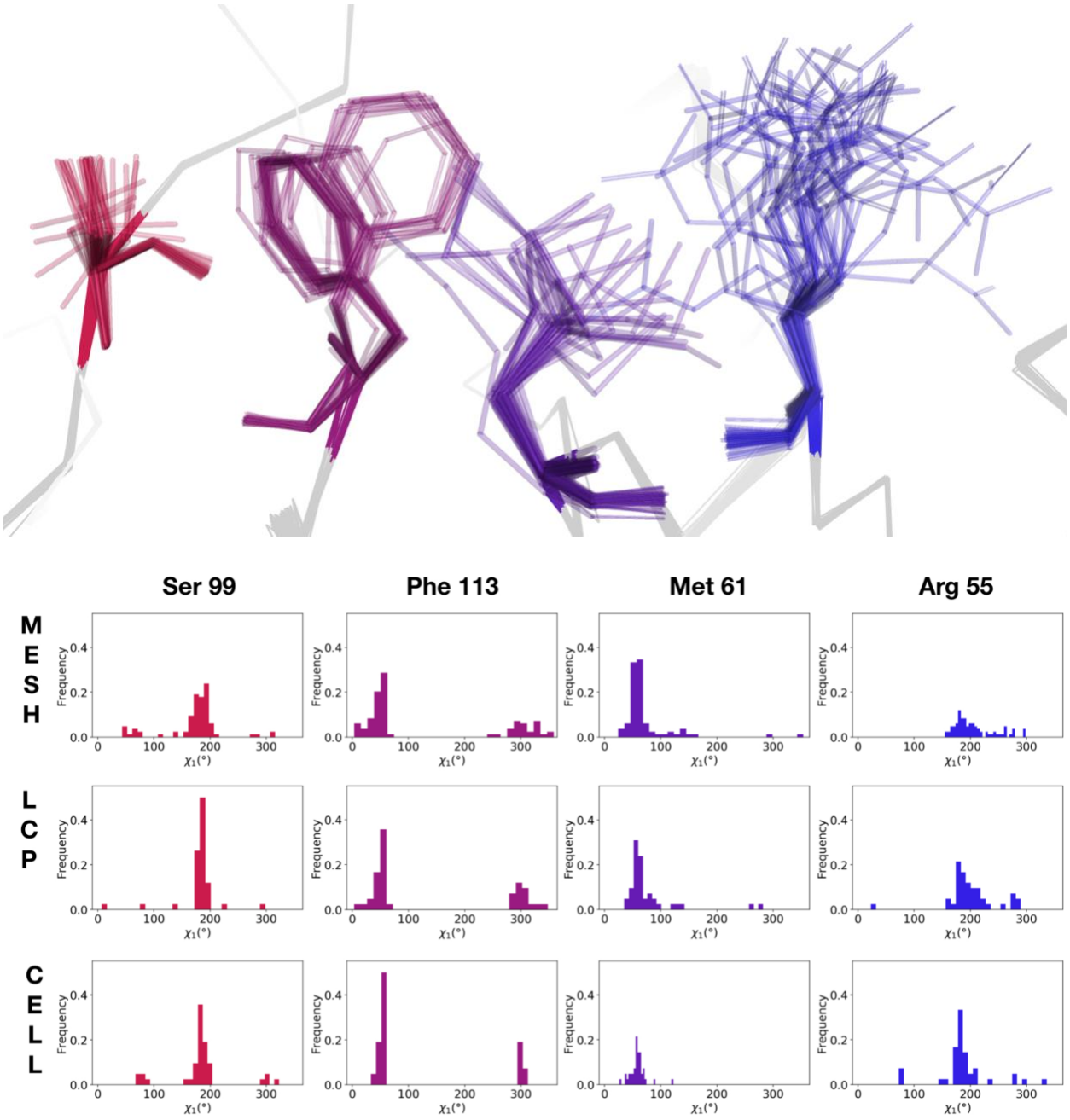

Figure 1.13. Visualization of the ensemble of conformers generated via phenix.ensemble_refine for the three serial XFEL datasets. The analysis focuses on a network of amino acid side chains that are known to be dynamic and important for function. In the top panel, the ensemble model for the cellulose dataset is displayed. Sticks are shown for the residues of interest (R55, M61, S99, F113), while the backbone is displayed as a ribbon for the rest of the structure. In the bottom panel, histograms of the distribution of chi1 angles are plotted for each of the four residues from the respective ensemble. Pairwise chi-square analysis revealed no significant differences between the rotamer distributions.

Table 1.6. Results of data indexing with XDS and unit cell refinement with REFMAC5.

Unit cell axis	a	b	c
XDS indexed (merged)	42.40	53.40	87.76
XDS indexed (std. dev.)	0.05	0.03	0.06
REFMAC5 refined	42.56	53.56	88.07

Table 1.7. Quantitative comparison of unit cell distributions across XFEL datasets. One-way ANOVA with post-hoc Tukey HSD tests revealed that all three datasets were significantly different across all three edges of the crystalline unit cell, as denoted by the respective *p*-values for each test.

	A axis	B axis	C axis
MESH Unit Cell (N = 18358)	43.32±0.11 Å	52.94±0.09 Å	89.87±0.21 Å
LCP Unit Cell (N = 11821)	43.10±0.18 Å	52.65±0.13 Å	89.29±0.26 Å
Cellulose Unit Cell (N = 23947)	43.00±0.26 Å	52.60±0.23 Å	89.20±0.37 Å
ANOVA	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$
MESH vs. LCP Tukey HSD Post-hoc Test	Difference = -0.22 95 % CI = -0.2256 to -0.2144 $p < 0.0001$	Difference = -0.29 95 % CI = -0.2948 to -0.2852 $p < 0.0001$	Difference = -0.58 95 % CI = -0.5883 to -0.5717 $p < 0.0001$
MESH vs. Cellulose Tukey HSD Post-hoc Test	Difference = -0.32 95 % CI = -0.3247 to -0.3153 $p < 0.0001$	Difference = -0.34 95 % CI = -0.3440 to -0.3360 $p < 0.0001$	Difference = -0.67 95 % CI = -0.6769 to -0.6631 $p < 0.0001$
LCP vs. Cellulose Tukey HSD Post-hoc Test	Difference = -0.10 95 % CI = -0.1053 to -0.0947 $p < 0.0001$	Difference = -0.05 95 % CI = -0.0545 to -0.0455 $p < 0.0001$	Difference = -0.09 95 % CI = -0.0979 to -0.0821 $p < 0.0001$

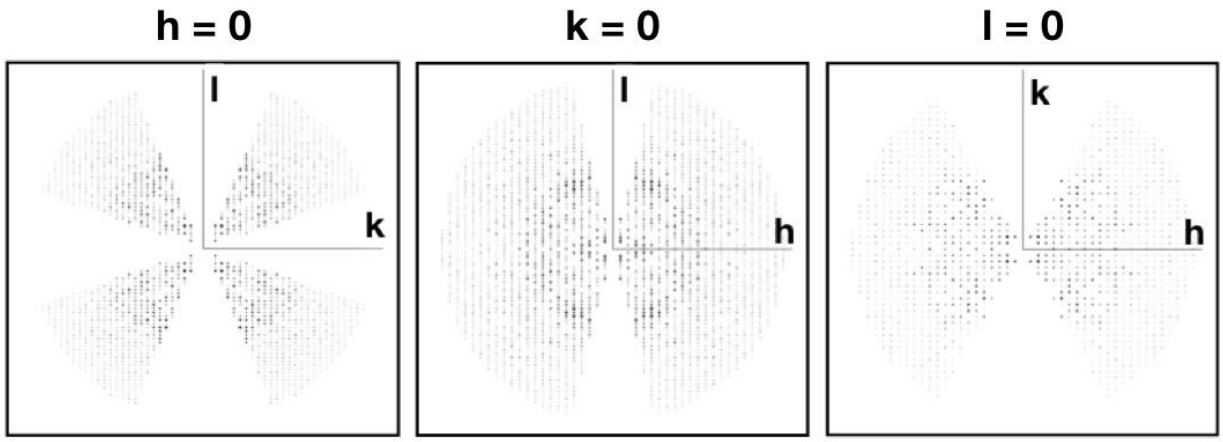

Figure 1.14. MicroED data visualized on two-dimensional slices of the reciprocal lattice. Measured reflections are visualized in black. Missing measurements along the kl plane may contribute to challenges in assignment of unit cell dimensions.

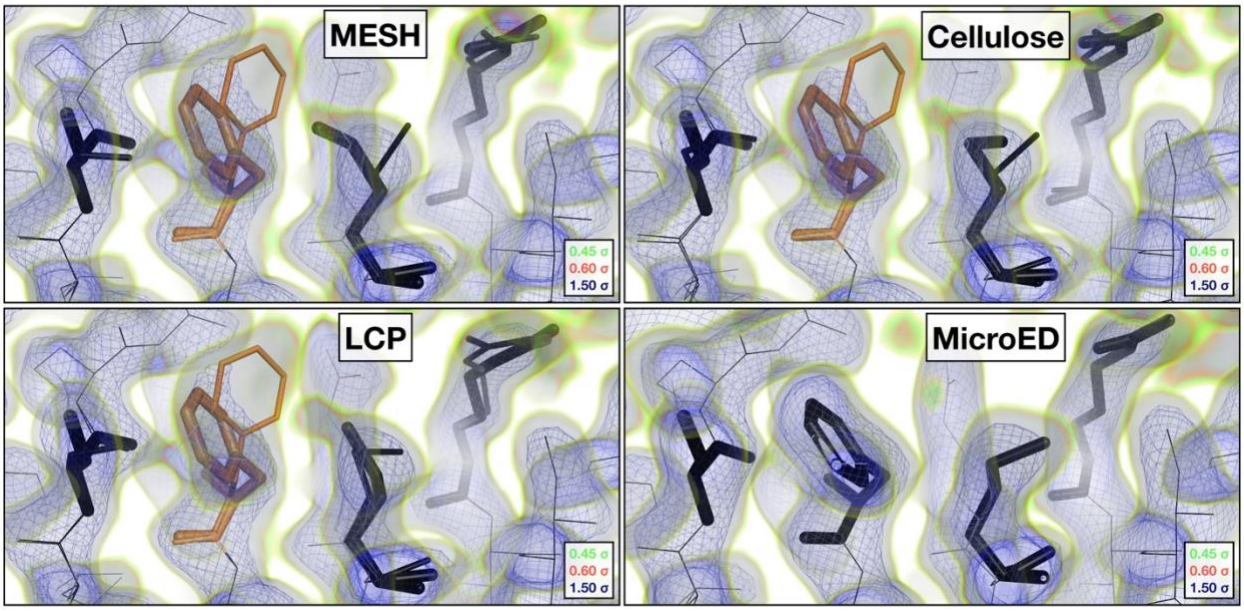

Figure 1.15. Visualization of all four datasets truncated at 2.5 Å. Comparison of the 2mFoFc maps and refined multi-conformer models reveals evidence for alternative conformations in the room temperature data (MESH, LCP, Cellulose), while cryogenic data (MicroED) supports a single conformer model.

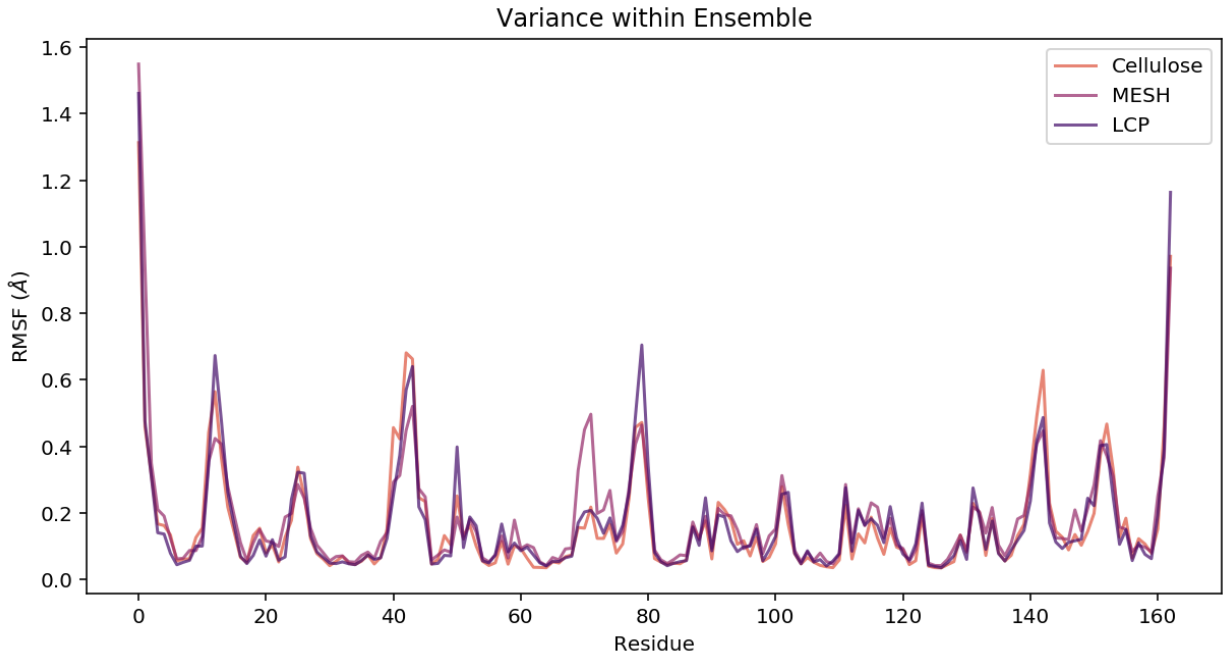

Figure 1.16. Average RMSF per residue for each ensemble generated for the serial XFEL datasets. A loop containing residues 64-74 samples more conformations in the MESH ensemble than in the other two.

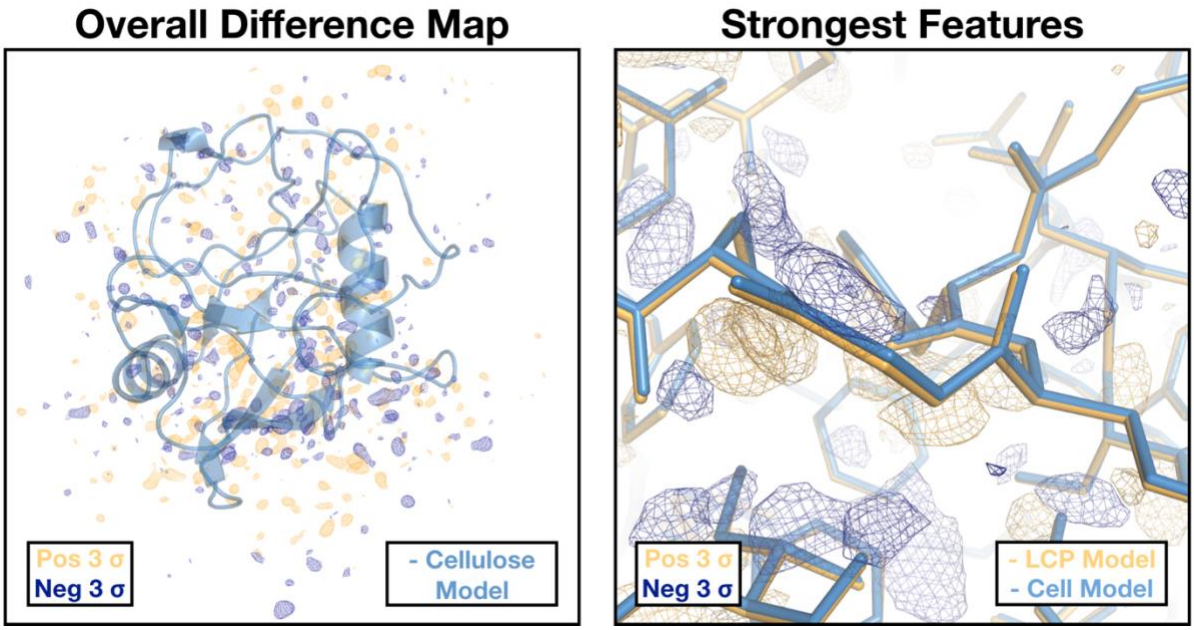

Figure 1.17. Visualization of an Fo-Fo difference map, displayed at 3 sigma, for the LCP and Cellulose datasets. In the left panel the overall difference map is shown, with the Cellulose model visualized for structural context. In the right panel the strongest features within the map are visualized, with both the LCP model and the Cellulose (Cell) model visualized for comparison.

References

1. Cusack, S. *et al.* Small is beautiful: protein micro-crystallography. *Nat. Struct. Mol. Biol.* **5**, 634–637 (1998).
2. Zander, U. *et al.* MeshAndCollect: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 2328–2343 (2015).
3. Liu, W. *et al.* Serial Femtosecond Crystallography of G Protein–Coupled Receptors. *Science* **342**, 1521–1524 (2013).
4. Zhou, X. E. *et al.* X-ray laser diffraction for structure determination of the rhodopsin-arrestin complex. *Sci. Data* **3**, 160021 (2016).
5. Chernov, A. A. Estimates of internal stress and related mosaicity in solution grown crystals: proteins. *J. Cryst. Growth* **196**, 524–534 (1999).
6. Coquelle, N. *et al.* Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* **10**, 31–37 (2018).
7. Olmos, J. L. *et al.* Enzyme intermediates captured “on the fly” by mix-and-inject serial crystallography. *BMC Biol.* **16**, 59 (2018).
8. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011).
9. Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods* **11**, 927–930 (2014).
10. Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *eLife* **2**, e01345 (2013).
11. Nannenga, B. L. & Gonen, T. The cryo-EM method microcrystal electron diffraction (MicroED). *Nat. Methods* **16**, 369–379 (2019).

12. de la Cruz, M. J. *et al.* Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **14**, 399–402 (2017).
13. Young, I. D. *et al.* Structure of photosystem II and substrate binding at room temperature. *Nature* **540**, 453–457 (2016).
14. Nango, E. *et al.* A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* **354**, 1552–1557 (2016).
15. Baxter, E. L. *et al.* High-density grids for efficient data collection from multiple crystals. *Acta Crystallogr. Sect. Struct. Biol.* **72**, 2–11 (2016).
16. Mueller, C. *et al.* Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography. *Struct. Dyn.* **2**, 054302 (2015).
17. Fuller, F. D. *et al.* Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat. Methods* **14**, 443–449 (2017).
18. Hunter, M. S. *et al.* Fixed-target protein serial microcrystallography with an x-ray free electron laser. *Sci. Rep.* **4**, 6026 (2014).
19. Sierra, R. G. *et al.* Nanoflow electrospinning serial femtosecond crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1584–1587 (2012).
20. Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 3309 (2014).
21. Sugahara, M. *et al.* Hydroxyethyl cellulose matrix applied to serial crystallography. *Sci. Rep.* **7**, 703 (2017).
22. Rodriguez, J. A. *et al.* Structure of the toxic core of α-synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).
23. Duyvesteyn, H. M. E. *et al.* Machining protein microcrystals for structure determination by electron diffraction. *Proc. Natl. Acad. Sci.* (2018) doi:10.1073/pnas.1809978115.
24. Martynowycz, M. W., Zhao, W., Hattne, J., Jensen, G. J. & Gonen, T. Collection of Continuous Rotation MicroED Data from Ion Beam-Milled Crystals of Any Size. *Structure*

27, 545-548.e2 (2019).

25. Martynowycz, M. W., Zhao, W., Hattne, J., Jensen, G. J. & Gonen, T. Qualitative Analyses of Polishing and Precoating FIB Milled Crystals for MicroED. *Structure* **27**, 1594-1600.e2 (2019).
26. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
27. Keedy, D. A. *et al.* Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife* **4**, e07574 (2015).
28. Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).
29. Ishchenko, A., Cherezov, V. & Liu, W. Preparation and Delivery of Protein Microcrystals in Lipidic Cubic Phase for Serial Femtosecond Crystallography. *J. Vis. Exp. JoVE* (2016) doi:10.3791/54463.
30. Sierra, R. G. *et al.* The Macromolecular Femtosecond Crystallography Instrument at the Linac Coherent Light Source. *J. Synchrotron Radiat.* **26**, 346–357 (2019).
31. Sierra, R. G. *et al.* Concentric-flow electrokinetic injector enables serial crystallography of ribosome and photosystem II. *Nat. Methods* **13**, 59–62 (2016).
32. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 85–97 (2018).
33. Brewster, A. S. *et al.* Improving signal strength in serial crystallography with DIALS geometry refinement. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 877–894 (2018).
34. Sauter, N. K. XFEL diffraction: developing processing methods to optimize data quality. *J. Synchrotron Radiat.* **22**, 239–248 (2015).
35. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292 (2011).
36. Ishikawa, T. *et al.* A compact X-ray free-electron laser emitting in the sub-ångström region.

Nat. Photonics **6**, 540–544 (2012).

37. Tono, K. *et al.* Diverse application platform for hard X-ray diffraction in SACLA (DAPHNIS): application to serial protein crystallography using an X-ray free-electron laser. *J. Synchrotron Radiat.* **22**, 532–537 (2015).
38. Tono, K., Hara, T., Yabashi, M. & Tanaka, H. Multiple-beamline operation of SACLA. *J. Synchrotron Radiat.* **26**, 595–602 (2019).
39. Kameshima, T. *et al.* Development of an X-ray pixel detector with multi-port charge-coupled device for X-ray free-electron laser experiments. *Rev. Sci. Instrum.* **85**, 033110 (2014).
40. Nakane, T. *et al.* Data processing pipeline for serial femtosecond crystallography at SACLA. *J. Appl. Crystallogr.* **49**, 1035–1041 (2016).
41. Barty, A. *et al.* Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Crystallogr.* **47**, 1118–1131 (2014).
42. White, T. A. *et al.* Recent developments in CrystFEL. *J. Appl. Crystallogr.* **49**, 680–689 (2016).
43. Duisenberg, A. J. M. Indexing in single-crystal diffractometry with an obstinate list of reflections. *J. Appl. Crystallogr.* **25**, 92–96 (1992).
44. Kabsch, W. *XDS*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
45. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
46. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
47. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
49. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.0.

50. Burnley, B. T., Afonine, P. V., Adams, P. D. & Gros, P. Modelling dynamics in protein crystal structures by ensemble refinement. *eLife* **1**, e00311 (2012).
51. Ibrahim, M. *et al.* Improvements in serial femtosecond crystallography of photosystem II by optimizing crystal uniformity using microseeding procedures. *Struct. Dyn.* **2**, 041705 (2015).
52. Sugahara, M. *et al.* Grease matrix as a versatile carrier of proteins for serial crystallography. *Nat. Methods* **12**, 61–63 (2015).
53. Suga, M. *et al.* Light-induced structural changes and the site of O=O bond formation in PSII caught by XFEL. *Nature* **543**, 131 (2017).
54. Kubo, M. *et al.* Nanosecond pump–probe device for time-resolved serial femtosecond crystallography developed at SACLA. *J. Synchrotron Radiat.* **24**, 1086–1091 (2017).
55. Martin-Garcia, J. M. *et al.* Serial millisecond crystallography of membrane and soluble protein microcrystals using synchrotron radiation. *IUCrJ* **4**, 439–454 (2017).
56. Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121 (2005).
57. Clabbers, M. T. B. & Abrahams, J. P. Electron diffraction and three-dimensional crystallography for structural biology. *Crystallogr. Rev.* **24**, 176–204 (2018).
58. Clabbers, M. T. B., Gruene, T., Parkhurst, J. M., Abrahams, J. P. & Waterman, D. G. Electron diffraction data processing with DIALS. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 506–518 (2018).
59. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
60. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
61. Baitan, D. *et al.* Growing Protein Crystals with Distinct Dimensions Using Automated Crystallization Coupled with In Situ Dynamic Light Scattering. *J. Vis. Exp. JoVE* (2018)

doi:10.3791/57070.

62. Schubert, R. *et al.* Real-Time Observation of Protein Dense Liquid Cluster Evolution during Nucleation in Protein Crystallization. *Cryst. Growth Des.* **17**, 954–958 (2017).
63. Caines, M. E. C. *et al.* Diverse HIV viruses are targeted by a conformationally dynamic antiviral. *Nat. Struct. Mol. Biol.* **19**, 411–416 (2012).
64. Hekstra, D. R. *et al.* Electric-field-stimulated protein mechanics. *Nature* **540**, 400–405 (2016).
65. Zatsepin, N. A., Li, C., Colasurd, P. & Nannenga, B. L. The complementarity of serial femtosecond crystallography and MicroED for structure determination from microcrystals. *Curr. Opin. Struct. Biol.* (2019) doi:10.1016/j.sbi.2019.06.004.
66. Halle, B. Biomolecular cryocrystallography: Structural changes during flash-cooling. *Proc. Natl. Acad. Sci.* **101**, 4793–4798 (2004).
67. Hattne, J., Shi, D., Cruz, M. J. de la, Reyes, F. E. & Gonen, T. Modeling truncated pixel values of faint reflections in MicroED images. *J. Appl. Crystallogr.* **49**, 1029–1034 (2016).

Chapter II

Temperature-jump solution X-ray scattering reveals distinct motions in a dynamic enzyme

Contributing Authors

Michael C. Thompson [1], Benjamin A. Barad [1,2], Alexander M. Wolff [1,2], Hyun Sun Cho [3], Friedrich Schotte [3], Daniel M. C. Schwarz [1,4], Philip Anfinrud [3] and James S. Fraser [1]

1. Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA.
2. Biophysics Graduate Program, University of California, San Francisco, San Francisco, CA, USA.
3. Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA.
4. Chemistry and Chemical Biology Graduate Program, University of California, San Francisco, San Francisco, CA, USA.

Preface

The bulk of this work appears as Thompson *et al.*, published in *Nature Chemistry* in 2019.

When I joined the Fraser Lab I wanted to work on time-resolved temperature-jump investigations of protein dynamics. Michael Thompson was the leader of the temperature-jump project, and we hoped to conduct our experiments in the context of crystalline diffraction. Nonetheless, Michael and Ben Barad were also conducting temperature-jump solution scattering experiments as a proof of concept. Michael provided the vision and leadership, while Ben spearheaded the data analysis. During this work, Ben taught me how to carry out real-time analysis using Python. This feedback was critical because we needed to assess the temperature-jump during our experiment. Initially, we asked the qualitative question, “are we getting a temperature-jump?” Next, we asked, “what is the magnitude of the temperature-jump?”

As a result, I became ever more involved in the temperature-jump project, and had the opportunity to contribute to the work described in the following chapter.

Abstract

Correlated motions of proteins are critical to function, but these features are difficult to resolve using traditional structure determination techniques. Time-resolved X-ray methods hold promise for addressing this challenge, but have relied on the exploitation of exotic protein photoactivity, and are therefore not generalizable. Temperature jumps, through thermal excitation of the solvent, have been utilized to study protein dynamics using spectroscopic techniques, but their implementation in X-ray scattering experiments has been limited. Here, we perform temperature-jump small- and wide-angle X-ray scattering measurements on a dynamic enzyme, cyclophilin A, demonstrating that these experiments are able to capture functional intramolecular protein dynamics on the microsecond timescale. We show that cyclophilin A displays rich dynamics following a temperature jump, and use the resulting time-resolved signal to assess the kinetics of conformational changes. Two relaxation processes are resolved: a fast process is related to surface loop motions, and a slower process is related to motions in the core of the protein that are critical for catalytic turnover.

Introduction

Protein motions are critical for functions such as enzyme catalysis and allosteric signal transduction[1], but it remains challenging to study excursions away from the most populated conformations[2]. Traditional methods that utilize X-rays for structural characterization of biological macromolecules, such as crystallography and solution scattering, provide high-quality structural information, but this information is both spatially and temporally averaged because the measurements are performed on large ensembles of molecules and are typically slower than the timescales of molecular motion[2,3]. To some extent, the spatial averaging inherent to X-ray

experiments is advantageous, because it reveals the alternative local conformations of a molecule that are significantly populated at equilibrium; however, structural states that are not significantly populated at equilibrium, such as intermediates along a conformational transition pathway, are effectively invisible. The temporal averaging inherent to X-ray experiments also results in a loss of information about how transitions between local alternative conformational states are coupled to one another. To gain kinetic information about molecular motion, researchers often turn to spectroscopic methods, but it can be difficult to correlate spectroscopic observables with high-resolution structural models.

Time-resolved X-ray scattering and diffraction can overcome the limitations of traditional structure determination for studying the dynamics of biomolecules[4–7]. In these experiments, a fast perturbation is applied to the sample to remove it from conformational equilibrium and synchronize conformational changes in a significant fraction of the molecules. Ultrafast X-ray pulses, which are short relative to motions of interest, are then used to perform structural measurements in real time as the system relaxes to a new equilibrium, providing simultaneous structural and kinetic information at high spatial and temporal resolution. Time-resolved X-ray experiments can identify transiently populated structural states along a conformational transition pathway, and reveal kinetic couplings between conformations[8]. Despite this potential to provide a wealth of information, especially when combined with molecular dynamics simulation[9–12], time-resolved experiments have not been broadly applied by structural biologists. To date, systems that have been most rigorously studied are those in which a protein conformational change is coupled to excitation of a photoactive chromophore molecule, because the conformational change can be initiated with an ultrafast optical laser pulse (for example, refs. [13–18]). Unfortunately, the number of proteins that undergo specific photochemistry as part of their functional cycle is small, and there is a fundamental need to develop generalized methods that can be used to synchronously excite conformational transitions in any protein molecule and expand the utility of time-resolved structural experiments[19–27].

Protein structural dynamics are intimately coupled to the thermal fluctuations of the surrounding solvent ('solvent slaving'[28,29]), and thermal excitation of the solvent by infrared laser temperature jump has been used in numerous pump–probe experiments. These experiments work on the principle that absorption of infrared photons excites the O–H stretching modes of water molecules, and the increased vibrational energy is dissipated through increased rotation and translation of the solvent molecules, effectively converting electromagnetic energy into kinetic (thermal) energy. Because this process of solvent heating and subsequent heat transfer to the protein is much faster than the large-scale molecular motions that define protein conformational changes, the sudden temperature jump removes conformational ensembles of protein molecules from their thermal equilibrium so that their structural dynamics can be measured using relaxation methods (**Figure 2.1a**). For example, temperature-jump perturbations have been coupled to ultrafast spectroscopic methods, including Fourier-transform infrared spectroscopy[30,31], NMR[32–34] and various forms of fluorescence spectroscopy[35,36], for the study of protein folding and enzyme dynamics. While these methods provide detailed kinetic information, they yield only very limited structural information about the underlying atomic ensemble. In contrast, the application of temperature jumps to time-resolved X-ray scattering and diffraction has been very limited. Nearly two decades ago, Hori et al.[37] used temperature-jump Laue crystallography to study the initial unfolding step of 3-isopropylmalate dehydrogenase. That study explored only a single pump–probe time delay, which allowed them to observe laser-induced structural changes but precluded kinetic analysis. Within the past two years, the laser temperature-jump method has been paired with X-ray solution scattering to explore the oligomerization of insulin under non-physiological conditions[25,26] and haemoglobin[27]. The results and analysis we present here expand the role of the temperature-jump method in structural biology, by showing that temperature-jump X-ray scattering experiments can be used as a general method to explore the functional, internal dynamics of proteins in solution. Additionally, we provide a detailed outline of a data reduction and analysis procedure suitable

for temperature-jump small-angle X-ray scattering (SAXS)/wide-angle X-ray scattering (WAXS) experiments.

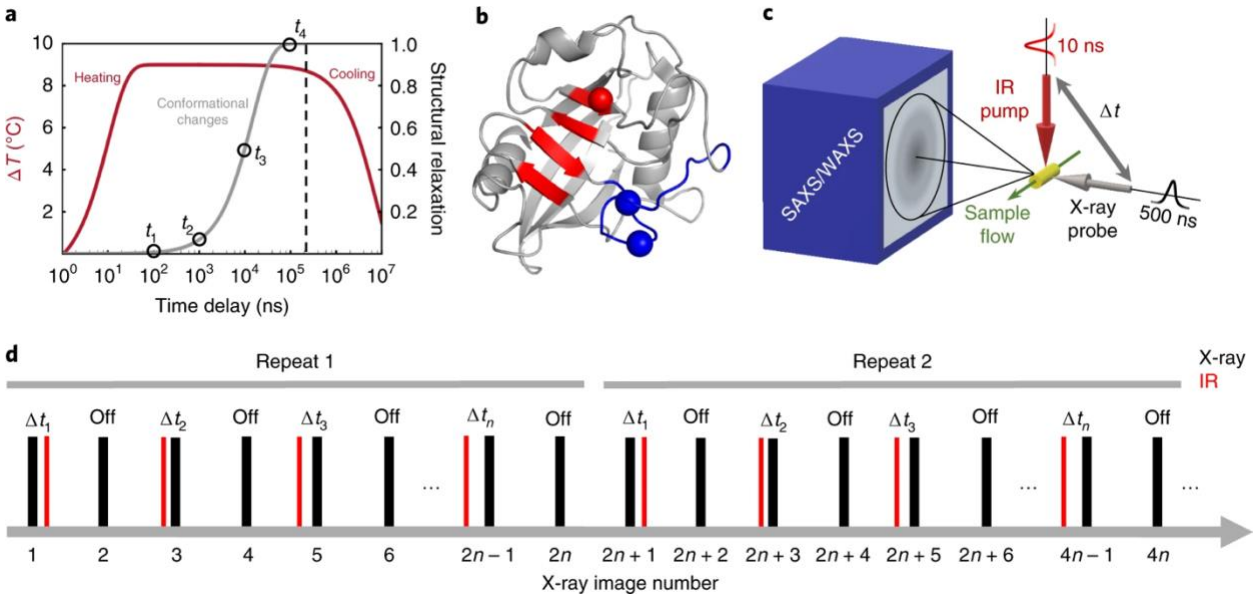

Figure 2.1. Diagram of temperature-jump experiment. a, During a temperature-jump experiment, an infrared (IR) laser pulse, several nanoseconds in duration, vibrationally excites the water O–H stretch and rapidly heats an aqueous solution of protein molecules (red curve). Heating is fast, but kinetic barriers to protein motions cause a lag in the structural relaxation to a new thermal equilibrium of conformational states (grey curve). **b,** Ribbon model depicting a single CypA molecule, with the 'core' dynamic residues (linked to catalysis) colored red and the 'loop' region adjacent to the active site (determines substrate specificity) colored blue. Sites of key mutations (S99T) are identified by spheres at the corresponding Cα positions (red: S99T; blue: D66N/R69H). **c**, Schematic of the temperature-jump SAXS/WAXS instrumentation, with key features highlighted. Liquid sample flows through the interaction region, where it interacts with mutually perpendicular IR pump and X-ray probe beams. Both pump and probe sources are pulsed, with a defined time delay between arrivals at the sample. SAXS/WAXS patterns are recorded on a single detector panel. **d,** Data collection sequence used for the temperature-jump experiments. For each pump–probe time delay, a pump–probe measurement ('laser on') was performed, followed by a measurement with no application of the pump laser ('laser off'). On–off pairs with increasing pump–probe time delays were measured in succession until all of the desired delay times were acquired, then the sequence was repeated as many as 50 times to improve the signal-to-noise ratio of the data. Note that the first measurement within each repeat is a control measurement, wherein the probe pulse arrives at the sample before the pump pulse (negative time delay).

The temperature-jump SAXS/WAXS experiments we describe here used human cyclophilin A (CypA)—a proline isomerase enzyme that functions as a protein-folding chaperone

and as a modulator of intracellular signaling pathways. CypA has been the subject of many NMR experiments that have identified two primary dynamic features of interest (**Figure 2.1b**). First, the active site-adjacent loops (covering approximately residues 60–80 and hereafter referred to as the 'loops' region) are mobile on the millisecond timescale[38]. This region is especially interesting because evolutionarily selected mutations within these loops perturb the dynamics of the loop[39], alter the binding specificity of CypA for substrates such as immunodeficiency virus capsids[40], and restrict the host range of these viruses[41,42]. Second, a group of residues that extends from the active site into the core of the protein (hereafter referred to as the 'core' region) has also been shown to be mobile on a millisecond timescale[38]. Subsequent work incorporating multi-temperature X-ray crystallography[43], mutagenesis[44] and further NMR experiments[45] has established a relationship between catalysis and conformational dynamics of a group of side chains in this region. Motivated by the sensitivity of the conformational state of the active site–core network to temperature[43], we performed infrared laser-driven temperature jumps on buffered aqueous solutions of CypA and measured subsequent, time-dependent changes in SAXS/WAXS. While our measurements provide only low-resolution structural information, we were able to measure the kinetics of protein conformational changes in CypA. We identified two relaxation processes, and by performing temperature-jump experiments at a range of different temperatures, were able to calculate thermodynamic properties of the transition states for the underlying conformational transitions. Specific mutants in the 'loops' or the 'core' regions of CypA show that the two processes are independent, each representing a distinct and uncoupled reaction coordinate on a complex conformational landscape. Collectively, our measurements and analysis show that a wealth of information about a protein's conformational landscape can be obtained by pairing laser-induced temperature jump with time-resolved X-ray scattering.

Results

A method for simultaneous measurement of structural and kinetic details of intrinsic protein dynamics

To measure protein structural dynamics, we utilized a pump–probe method that pairs an infrared laser-induced temperature jump with global measurement of protein structure via X-ray solution scattering (**Figure 2.1c**). We performed solvent heating in aqueous protein solutions by exciting the water O–H stretch with mid-IR laser pulses (1,443 nm; 7 ns duration). At regularly defined time delays following the infrared heating pulse (from 562 ns to 1 ms), we probed the sample with high-brilliance synchrotron X-ray pulses from a pink-beam undulator (3% bandwidth at 12 keV; **Figure 2.6**) that were approximately 500 ns in duration, and measured X-ray scattering using a large charge-coupled device detector that was capable of capturing small and wide scattering angles on a single panel. Because the duration of the infrared pump pulse was sufficiently short compared with the duration of the X-ray probe, the heating was effectively instantaneous with respect to the relaxation processes we were able to observe. Data were collected as interleaved 'laser on' and 'laser off' X-ray scattering images, so that each pump–probe measurement could be paired to a measurement made immediately before application of the pump laser (**Figure 2.1d**). We measured 27 unique pump–probe time delays across four decades of time spanning from 562 ns to 1 ms, performing 50 repeat measurements for each time delay. For each detector image, the individual pixel values were azimuthally averaged as a function of the scattering vector magnitude, q, to give one-dimensional scattering intensity profiles (I(q) curves). All scattering profiles were scaled to a single reference, and the data were analyzed as described below. These pump–probe measurements allowed us to monitor structural changes within the ensemble of heated molecules in real time as the system relaxed to a new thermal equilibrium following temperature jump (**Figure 2.1a**). Additional details about the instrumentation used for our experiments are available in the Supplementary Methods.

Calibrating the magnitude of the temperature jump

Because the isothermal compressibility of liquid water is highly temperature dependent[46], X-ray scattering from the bulk solvent acts as an exquisitely sensitive thermometer that can be used to calibrate the magnitude of the temperature jump in our experiments[9,25,27,47,48]. Our instrument configuration allowed us to record low-angle protein scattering (SAXS) and high-angle solvent scattering (WAXS) on the same detector image, enabling simultaneous measurement of CypA structure and water temperature. A method based on singular value decomposition of a matrix constructed from the scattering curves provided a simple way to measure the magnitude of the laser-induced temperature jump (Supplementary Data). The singular value decomposition (SVD) analysis showed that the average temperature jump produced by our infrared heating pulse was approximately 10.7 °C, and allowed us to judge when cooling of the system became significant, so that we could identify the maximum pump–probe time delay that was valid for our relaxation analysis. Additionally, the SVD analysis revealed that following laser temperature jump, the solvent reaches a new thermal equilibrium faster than the measurement dead time of our experiment (562 ns). This observation is consistent with other work, in which changes to the structure of bulk solvent following laser temperature jump have been shown to equilibrate within roughly 200 ns[49].

Temperature jump produces changes in the X-ray scattering profile of CypA

To determine the effect of the temperature jump, we initially averaged all data for a given time delay (**Figure 2.7**), examined the scattering profiles for differences, and observed a small laser-induced change in the low-q region of the scattering profiles. Next, we exploited the interleaved data collection scheme (**Figure 2.1d**) to increase the sensitivity of the experiment. We calculated the ‘on–off difference’ between the paired ‘laser on’ and ‘laser off’ scattering profiles. Following subtraction, we binned the ‘on–off difference’ scattering curves according to the associated pump–probe time delay, performed an iterative chi-squared test to remove outliers, and averaged the calculated differences for all repeat measurements (Supplementary

Methods). This subtraction and averaging removed systematic errors that accumulated over long experiments, resulting in accurate measurements of difference signals as a function of the pump–probe time delay. Scattering differences at high q (1.0–4.2 Å−1) were used to calibrate the final sample temperature after laser illumination (Supplementary Data), while differences at low q (0.03–0.3 Å−1) were analyzed in the context of the average physical dimensions and scattering density of the CypA 'protein particle'. Here, we use the phrase 'protein particle' to describe the protein molecule plus the ordered solvent bound to its surface, since both the protein and its hydration layer have electron densities that differ from bulk solvent, and therefore contribute to the observed X-ray scattering by the CypA solution. Identical temperature-jump measurements were performed on protein samples and on samples consisting of buffer only without protein, and an additional step of scaling and subtraction isolated the signal changes at low q due only to the protein (**Figure 2.7**).

Time-resolved changes in small-angle X-ray scattering

Comparison of difference scattering curves calculated for 27 unique time delays revealed a time-dependent change in X-ray scattering by the protein, demonstrating that the modest temperature jump we introduced was capable of exciting protein dynamics that could be observed in real time. The difference curves calculated from our data, showing the contribution of the protein to time-resolved changes in the SAXS/WAXS signal, have features in the low-q (q = 0.03–0.3 Å−1) region (**Figure 2.2a**). The time-resolved differences are approximately the same in magnitude and direction as differences between static temperature measurements performed on samples equilibrated to temperatures that differ by 10 °C, and correspond to the 'laser on' and 'laser off' temperatures (**Figure 2.8**). Qualitatively, the time-resolved on–off differences show that the overall low-angle scattering and extrapolated value of I(0) are reduced within the measurement dead time of our experiment (562 ns), then begin to increase slightly over the next few microseconds before decreasing further at longer pump–probe time delays out to 562 μs. The observed 'laser on–off' difference in I(0) is approximately 3% of the total

observed signal, with one-third of that signal change occurring in the time regime that can be resolved by our measurements. Additional static measurements as a function of both CypA concentration and temperature allowed us to characterize and correct for the effect of interparticle interactions, and to determine that their contribution to the X-ray scattering is temperature independent over the relevant temperature range (Supplementary Data).

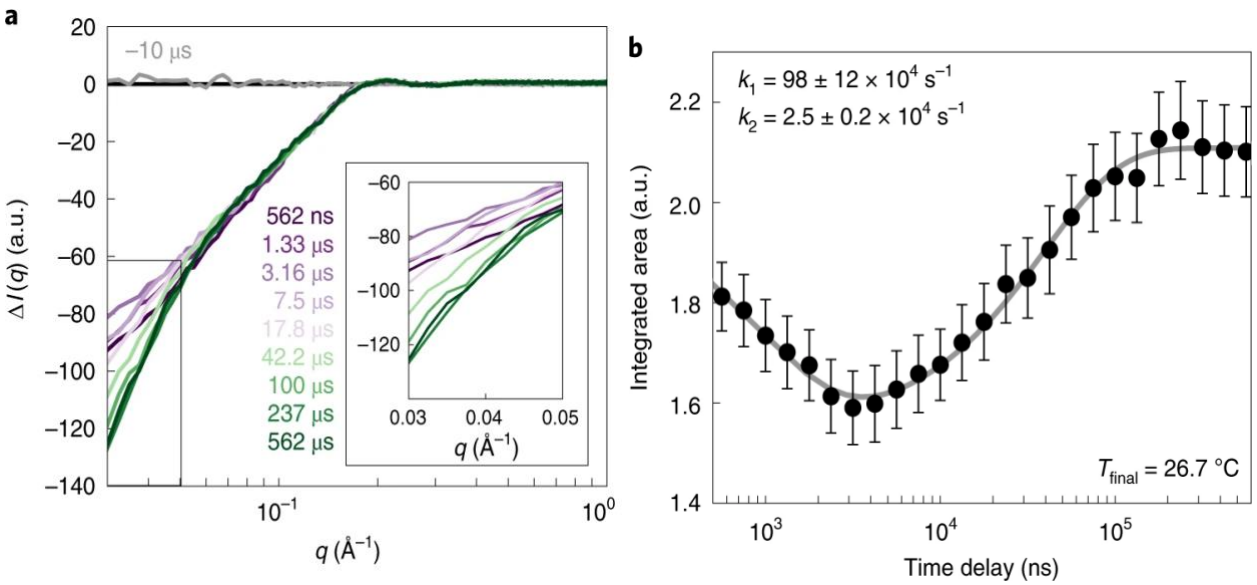

Figure 2.2. Time-resolved signal. a, A series of averaged time-resolved difference X-ray scattering curves is shown for a subset of our temperature-jump data (10 out of 26 unique time delays). Data at low q are plotted on a linear q scale in the inset. **b,** The area under the averaged difference scattering curve in the q = 0.03–0.05 Å−1 region was integrated for all measured pump–probe time delays, and the resulting absolute values and their standard deviations (error bars) are plotted as a function of the pump–probe time delay. The plotted data suggest the existence of multiple relaxation processes, and we used a two-step model of relaxation kinetics to fit the observations (grey line). The rates calculated from the kinetic fit are provided, along with their standard deviations.

Changes in low-angle scattering generally reflect changes in the overall size and shape of the particles in solution, and changes in I(0) result from changes to the scattering density of the particle (that is, the number of excess electrons in the particle relative to the bulk solvent that it displaces). The observed reduction of both quantities indicated shrinkage of the scattering particles, and a decrease in their scattering density. The on–off difference for our shortest

pump–probe time delay (562 ns) is significantly different from 0 at low scattering angles (a reduction in I(0) of approximately 2%), showing the existence of structural changes that are faster than the measurement dead time of our experiments. We hypothesize that the physical basis for this fast signal change is a combination of thermal disorder and thermal expansion effects (Supplementary Data), possibly including a temperature-induced loss (or 'melting') of ordered solvent within the protein hydration shell. Previous work, such as measurement of 'protein-quake' motions in photoactive systems[9,50], has shown that vibrational energy transfer is fast in proteins, and we expect that the onset of thermal disorder following temperature jump happens on a similar timescale, within hundreds of picoseconds. Furthermore, thermal expansion of solvent following temperature jump is well known to equilibrate within approximately 200 ns[25,47,48], and it is reasonable to assume that protein thermal expansion may occur on a similar timescale. The notion that thermal disorder and expansion occur rapidly in our system is experimentally supported by Kratky plots created from our static and temperature-jump data, which suggest a slight increase in protein flexibility without unfolding (**Figure 2.9**). The effect is temperature dependent, but not time dependent over the pump–probe time delays we explored, confirming that the process is faster than the measurement dead time of our experiment. The physical details and kinetics of these fast processes, although interesting, are invisible to our experiment, and it is impossible to disentangle and quantify the exact contribution of each phenomenon to the fast signal decrease that we observe. Therefore, our analysis focused on structural dynamics that occur in the microsecond regime, which is resolved by our measurements. In contrast, ultrafast single pulses from synchrotrons (pulse duration: ~100 ps) or X-ray free-electron lasers (pulse duration: ~10 fs) offer the opportunity to study fast timescale processes using temperature jump; however, a major consideration for temperature-jump SAXS/WAXS experiments on faster timescales is the additional complexity in signal interpretation resulting from the complex dynamics of the solvent that immediately follow laser heating[25,47,48] but equilibrate within the dead time of the experiments reported here.

Kinetic modelling of structural dynamics from time-resolved scattering differences

Our time-resolved measurements of scattering differences allowed us to model the kinetics of global structural changes induced by the temperature jump (Supplementary Methods). We integrated the area under each of our time-resolved difference curves in the q = 0.03–0.05 Å−1 region and plotted the absolute value of the area as a function of the associated pump–probe time delay (**Figure 2.2b**). Based on the apparent shape of the area versus time delay plot, we reasoned that a two-step kinetic model would be needed to fit the data, since the area first decreases and then increases as a function of time delay. We fit the observed data to a two-step model of relaxation kinetics (independent steps) using a nonlinear least-squares curve-fitting algorithm, and calculated rates of 98 × 104 ± 12 × 104 s−1 for the fast process (k1) and 2.5 × 104 ± 0.2 × 104 s−1 for the slow process (k2) at 26.7 °C (299.7 K). The errors calculated for these rates were determined using a bootstrap analysis[51]. We note that while other experimenters have employed SVD for kinetic analysis of time-resolved scattering, in our case, implementation of this method produced comparable results to the integration analysis, as one singular vector dominated the time-dependent signal identified by the SVD (**Figure 2.10**).

In addition to performing kinetic analysis, we also utilized the on–off difference curves to generate I(q)t scattering curves, which represent the time-dependent X-ray scattering from the CypA sample, but are generated in a manner that makes use of the paired ‘laser on’ and ‘laser off’ measurements to reduce the effects of systematic error (Supplementary Methods). We subsequently used these scattering curves for Guinier analysis (Supplementary Methods) to determine how the radius of gyration (Rg) of the average CypA particle in the conformational ensemble changes as a function of time following the temperature jump (**Figure 2.3**). This analysis showed that after the temperature jump, the average CypA particle shrinks in the dead time of our experiment; however, within a few microseconds of the temperature jump, a fast structural transition (described by k1 in our kinetic analysis) causes the average CypA particle to

expand subtly. While the increase in the calculated radius of gyration is small relative to the error on the Guinier fit for any single data point, our conclusion that the particle is expanding is supported by multiple time points and kinetic analysis of the integrated area under difference scattering curves. Following this fast increase in Rg, a second, slower process (described by k2 in our kinetic analysis) reverses this trend, causing the average CypA particle to shrink again.

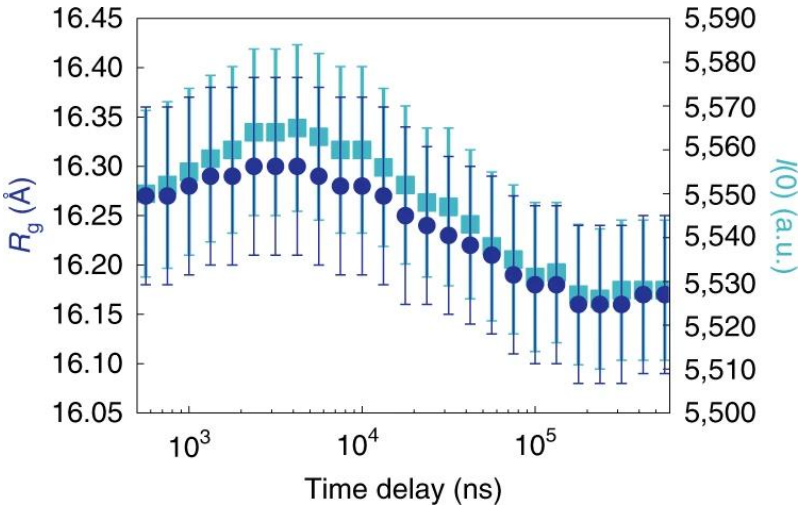

Figure 2.3. Changes in the average physical parameters of CypA protein particles in the ensemble were estimated by performing Guinier analysis on the time-resolved scattering data. Consistent with our kinetic analysis, the radius of gyration (Rg) of the average CypA particle in solution first increases and then decreases as a function of time following the temperature jump. Additionally, the value of I(0) extrapolated from the Guinier analysis shows an analogous increase and decrease, suggesting that the change in the particle size is coupled to a change in its average electron density, which is probably due to the acquisition and loss of water molecules from the solvation shell as the protein swells and then shrinks. Error bars represent standard deviations of the calculated quantities based on propagation of measurement standard deviations.

To learn more about the conformational transitions in CypA that are excited by the temperature jump, we repeated the experiment at multiple different jumped (final) temperatures ranging from 6.2–29.9 °C (279.2–302.9 K). We modelled the kinetics of the SAXS/WAXS signal changes to observe how the relaxation rates changed as a function of temperature. The calculated rates (k1 and k2) for all temperatures are provided in **Table 2.2**. We analyzed the temperature dependence of these rates, which provided insight into the thermodynamics of the

transition states for the two processes. We plotted ln[k/T] versus 1/T (**Figure 2.4**), and noted that for both processes the relationships appeared to be linear. Therefore, we calculated enthalpies and entropies of activation for each of the two processes by fitting the data to the linearized Eyring equation:

$$ln\left[\frac{k}{T}\right] = \left(\frac{\Delta H‡}{R} \times \frac{1}{T}\right) + \frac{\Delta S‡}{R} + ln\left[\frac{k_B}{h}\right] \quad \textbf{(Eq. 2.1)}$$

where R is the gas constant, kB is the Boltzmann constant, and h is Planck's constant. The enthalpies of activation (ΔH‡) and entropies of activation (ΔS‡) and their standard deviations are given in **Table 2.1**. The fast process (k1) has a large, positive enthalpy of activation, but this is partially offset by a slightly positive entropy of activation. Formation of the transition state during the slow process (k2) has a smaller enthalpic cost, but is also entropically disfavored. We note that the lowest temperature measurement (279.2 K) was not used in the Eyring analysis of the fast process (k1) because the error on the measured rate was large due to the low magnitude of the overall time-resolved signal changes at this temperature.

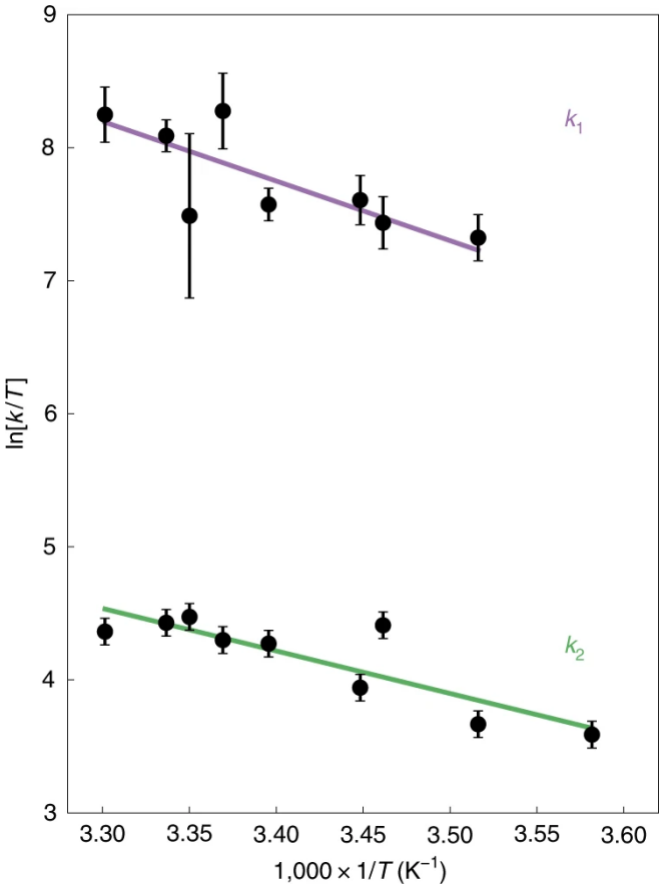

Figure 2.4. Linear Eyring plots are shown for each of the two relaxation processes observed in our temperature-jump experiments with CypA. Both processes show Arrhenius behavior over the temperature range we explored, and the apparent linear relationship between 1/T and ln[k/T] justifies the use of the linear Eyring equation (Eq. 2.1) to fit the data and characterize the thermodynamics of the transition states. Data points represent rates determined from the kinetic analysis, and error bars represent the standard deviations of the rates based on propagation of the measurement standard deviations and bootstrap analysis (see Supplementary Methods). The fit to the linear Eyring equation for the fast process (k1) is shown in purple, and the fit for the slow process (k2) is shown in green.

Table 2.1. Enthalpies (ΔH‡) and entropies (ΔS‡) of activation, and their standard deviations based on propagation of measurement standard deviations through all analysis steps, for the fast (k1) and slow (k2) processes observed for wild-type CypA, calculated from Eyring analysis.

	Δ*H*‡ (103 J mol−1)	Δ*S*‡ (J mol−1 K−1)
Fast process (*k* 1)	37 ± 8	32 ± 26
Slow process (*k* 2)	27 ± 3	−34 ± 11

CypA mutations with distinct effects on conformational dynamics alter time-resolved signal changes

The time-resolved signal changes that we attributed to wild-type CypA were observed only at low scattering angles; therefore, the resulting structural information had very limited resolution. To gain a better understanding of the structural transitions excited by the temperature jump, we next studied two specific CypA mutants, S99T (in the 'core' region; red in **Figure 2.1b**) and NH (D66N/R69H in the 'loops'; blue in **Figure 2.1b**). The conformational dynamics of these two variants of the enzyme each differ from the wild type in distinct ways: S99T is catalytically impaired due to a loss of rotameric exchange in a key network of residues, whereas NH alters the substrate specificity of CypA by enhancing the dynamics of the surface-exposed loops adjacent to the active site. Importantly, NMR relaxation measurements indicate that S99T perturbs the active site but not the loops[44], and that NH only perturbs the loops[39].

We observed that both S99T and NH mutants showed time-resolved SAXS signal changes that differed from the wild-type enzyme. Both mutants showed a fast signal change that occurred within the measurement dead time of the experiment, similar to what was observed for the wild-type enzyme, and consistent with these changes being largely due to thermal expansion or temperature-dependent changes to the solvation shell. Beyond the initial fast loss of scattering intensity that was observed (and nearly identical) for all three CypA variants we studied (wild type, S99T and NH), the evolution of the time-resolved signals for each of the two mutants differed substantially from the wild type and from one another. In the S99T mutant (**Figure 2.5a**), we observed only the fast decrease (k1) of the integrated area under the difference curve ($q = 0.03–0.05$ Å^{-1}), and a striking absence of the subsequent increase (k2) in the integrated area at longer time delays that was observed for the wild-type enzyme. In contrast, for NH, the plot of integrated area under the difference curve as a function of the time delay (**Figure 2.5b**) appears to lack the initial fast decrease (k1), but it does appear to retain the slower signal change (k2) that results in an increase for this quantity at longer time

delays. We found that for the mutants, the two-step kinetic model that was fit to the wild-type data yielded at least one rate with a large error. For the S99T mutant, the first step (k1) was well fit but the second step (k2) was poorly fit, while the opposite was true for the NH variant. After visual inspection, we chose to use a single-step kinetic model to fit the data for the S99T and NH mutants, and the calculated rates for the two mutants (k1 for S99T and k2 for NH) are also given in **Table 2.2**. Plots of the residuals for these fits revealed no structure, suggesting that a single-step kinetic model is sufficient to explain the data for the CypA mutants. In contrast, fitting kinetic data collected for the wild-type enzyme using a single-step model results in residuals with exponential character, and a two-step kinetic model is needed to reduce the error in the fit (**Figure 2.11**).

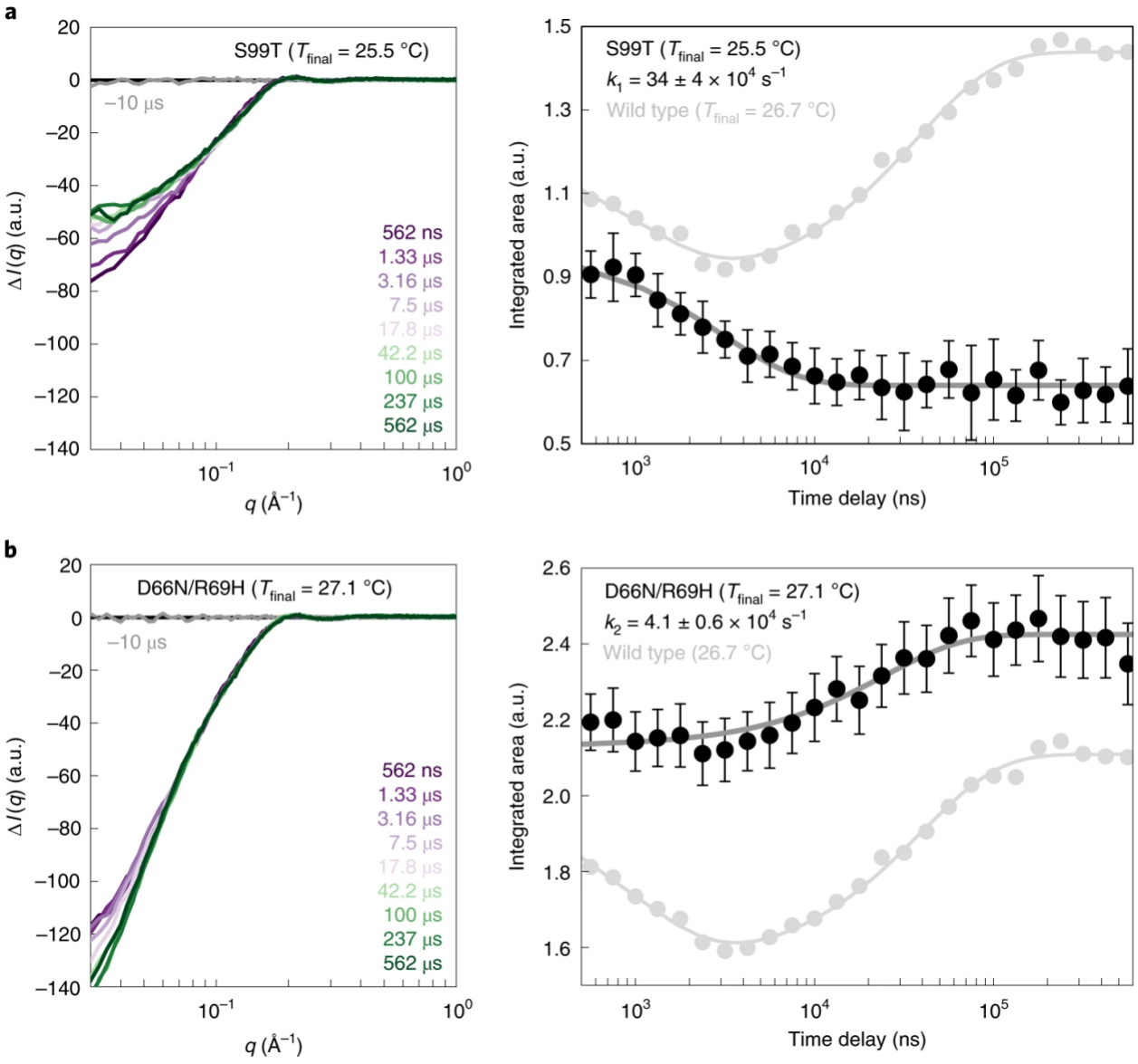

Figure 2.5. Effects of mutations upon time-resolved signal. Two CypA mutants (S99T (a) and D66N/R69H (b)), with distinct effects on the enzyme's function, show the link between the observed temperature-jump signal and functional dynamics. The data are presented in the same manner as for the wild-type enzyme shown in Figure 2.4. Left panels show a subset of averaged difference scattering curves for several pump–probe time delays. Data points in the right panels represent the absolute value of the integrated area under the corresponding averaged difference scattering curve in the q = 0.03–0.05 Å−1 region, and the error bars represent standard deviations. In the plots of integrated area versus pump–probe time delay (right panels), the signal observed for the wild-type enzyme is shown in light grey for comparison. In a, the S99T mutant, which displays defective catalytic function, shows only the fast relaxation process (k1) and lacks the slower process (k2). In b, the D66N/R69H (NH) mutant, with altered substrate specificity, shows the slow relaxation process (k2) and lacks the faster process (k1).

Our measurements of the S99T and NH variants of CypA clearly show that mutations that are known to impinge on the activity and specificity of the enzyme also perturb the observed time-resolved signal relative to the wild type in our temperature-jump experiments. Most notably, the slow relaxation process (modelled by k2) is shared only by the catalytically competent wild-type and NH variants, and its absence from the S99T variant suggests that the underlying conformational change is related to the catalytically coupled motions that are arrested by the S99T mutation. These results indicate that temperature-jump experiments are capable of exciting and measuring functionally relevant, intramolecular structural dynamics of proteins, even when the data are limited to relatively low scattering angles.

Discussion

Time-resolved temperature-jump X-ray scattering experiments have the potential to be a powerful tool for understanding the complex dynamics of protein molecules, such as the model enzyme CypA, which has no intrinsic photoactivity. In our experiments, temperature jumps of 10–11 °C modified the CypA conformational ensemble, producing a clear, time-dependent change at low scattering angles. Because the time-resolved scattering changes could be observed only at low q, we demonstrated that interparticle spacings (quantified by the structure factor, S(q)) were temperature independent, and that the changes were due to structural rearrangements within the protein particle. Guinier analysis of time-resolved scattering curves allowed us to track changes in the average radius of gyration (Rg) of the protein particles, which include the CypA molecules plus their ordered solvation shells. This signal change comprises an initial reduction in low-angle scattering that occurs within the measurement dead time of our experiment, followed by a small increase in low-angle scattering that equilibrates within a few microseconds ($k1 = 98 \pm 12 \times 104\ s^{-1}$ at 26.7 °C), and finally a further reduction in low-angle scattering that equilibrates within tens of microseconds ($k2 = 2.5 \pm 0.2 \times 104\ s^{-1}$ at 26.7 °C). High-angle scattering differences required for atomistic structural interpretation were not

observed due to signal-to-noise considerations. We suspect that high-angle features in time-resolved difference scattering curves may be especially weak for proteins such as CypA, in which conformational motions involve correlated shifts of atoms that can preserve many properties of short-range structure.

This analysis suggests a model in which the scattering density of the CypA particle (protein and ordered solvent) first increases and then decreases after excitation by the temperature jump. By performing temperature-jump experiments over a range of temperatures, we discovered that both the fast and slow processes we observed could be described using Arrhenius kinetics. An Eyring analysis revealed relatively large, positive enthalpies of activation for both processes, consistent with the idea that conformational changes generally require breakage of existing interactions in both the protein and the solvent. The activation enthalpy for the fast process (k1) is larger, but the overall activation energy is lower because of a favorable activation entropy. The opposite is true for the slower process (k2), which has a smaller overall activation enthalpy, but a disfavourable activation entropy. We used the S99T and NH variants of CypA to disentangle the nature of these processes and their associated functions. The S99T mutant is capable of undergoing the fast (k1) expansion process, but does not experience the subsequent slow (k2) shrinkage. NMR and crystallography have shown that this mutation arrests the conformational exchange of the 'core' catalytic network of residues in CypA by creating steric hindrance, strongly favouring a minor conformation of the wild-type enzyme[45]. This interpretation suggests that the internal rearrangements are related to the k2 process. However, there is a separation of timescales between the NMR results, which indicate millisecond dynamics in the 'core' region, and the temperature-jump SAXS results here, which indicate microsecond dynamics. This discrepancy may reflect coupled processes that are related by a population-shuffling mechanism[52] and agree with a broad timescale range of side chain dynamics in CypA uncovered by molecular dynamics experiments[53]. In contrast with the S99T mutant, the NH variant lacks the initial fast signal change (k1) in our temperature-jump

experiments, but clearly retains the slow (k2) signal. NMR and crystallographic studies of the D66N/R69H (NH) double mutant showed that it maintains wild-type catalytic motions, but breakage of several hydrogen-bonding interactions results in enhanced flexibility of a surface loop adjacent to the active site and altered substrate specificity[39]. NMR experiments with NH have shown that the loop motions still occur, but at an increased rate that renders them invisible to our experiments. Therefore, we hypothesize that the loop motions are responsible for the fast (k1) signal in wild-type CypA and S99T, where these motions have been shown (by NMR) to be unperturbed[44].

The assignment of the motions responsible for the experimentally observed Rg and I(0) changes by mutational analysis is also consistent with X-ray crystal structures of CypA. Using a room temperature X-ray crystal structure of wild-type CypA (Protein Data Bank ID: 3K0N), we calculated the Rg of the enzyme with the core catalytic network (Arg55, Met61, Ser99 and Phe113) in both the major and minor conformational states, and found that the predicted Rg of the minor state is 0.07 Å smaller than the major state (14.09 versus 14.16 Å, for protein atoms only). In addition to decreasing the Rg of the protein, conversion from the major to the minor state also results in a small reduction in solvent-exposed surface area, which would necessarily reduce the size of the protein solvation layer. While the reduction of the protein's Rg makes it more compact, and therefore should increase I(0) because the protein has become more dense, the loss of material from the solvation layer opposes this effect, probably leading to the observed decrease in the scattering density of the CypA particle and reduction in I(0). This coupling of Rg changes with changes in the solvent-exposed surface area of the protein can potentially explain the observed correlation between Rg and I(0) changes calculated from our time-resolved scattering data. Additionally, the increase in the average Rg and I(0) during the faster process (k1) is consistent with the loops sampling an expanded conformational ensemble (with increased surface area), as indicated by recent exact nuclear Overhauser effect NMR ensembles[54]. Our kinetic modelling of the wild-type and mutant data suggests that two dynamic

modes are observed, each of which is individually perturbed by different mutations. However, the extent to which the observed motions are coupled is unclear from our measurements. The S99T variant clearly shows that the fast motion can occur independent of the slower motion, because S99T is known to have arrested slow dynamics. In contrast, the fast motion we observe is likely to be accelerated by the NH mutation, becoming too fast for us to observe rather than being impeded. Therefore, we are unable to determine whether the fast motion is a requisite first step that precedes the slower motion.

Time-resolved X-ray structural measurements are critical for decoupling the experimental signatures of conformational changes that can become convoluted by the spatial and temporal averaging that is inherent to traditional X-ray experiments. If one were to assess traditional, static SAXS data for CypA, they would find that increasing the temperature of the sample results in a decrease in the average particle size at equilibrium. These static measurements as a function of equilibrium temperature fail to capture that the temperature change actually perturbs two distinct protein motions, which have the opposite effect on the enzyme's global structural characteristics. This information can only be obtained through a time-resolved experiment, which is able to separate the effects of these two motions because they have substantially different rates. The ability to dissect individual conformational motions and measure their rates is important for understanding processes involving complex protein dynamics. Many of these dynamic processes, including allostery[55–58] and enzyme catalysis[59–63], involve extensive reorganization of interactions between the protein and its ordered solvation shell, which are key contributors to the energetics that govern protein motions[28,29,64–68]. Because X-ray solution scattering experiments report on the structure of a protein and the ordered solvent molecules that constitute its solvation shell[69–72], the widespread application of time-resolved SAXS/WAXS experiments will enhance our understanding of protein motions, including how they are driven by solvent dynamics, especially when they can be combined with longer

and increasingly detailed molecular dynamics simulations to provide atomic-scale insight into the underlying structural changes[9–12].

For time-resolved experiments to enter the mainstream of structural biology, it is necessary to create general perturbations that can be applied universally (to any protein of interest), and our results establish that temperature jump can excite functional intramolecular protein motions for time-resolved X-ray structural measurements. The intrinsic motions of proteins and other biological macromolecules are naturally driven by thermal fluctuations, which makes temperature jump an ideal perturbation for studying their dynamics using time-resolved X-ray scattering. Consistent with this idea, a growing number of X-ray diffraction studies across multiple systems have shown that conformational equilibria, which are sensitive to temperature perturbation, are the same ones exploited by evolution to create new functions[73,74], by medicinal chemists in identifying novel small-molecule binding sites[75–77], and by enzymes during their catalytic cycles[44,78,79], paving the way for time-resolved measurements on a broad variety of biochemical systems. Temperature-jump solution scattering experiments are powerful tools for understanding protein dynamics, especially when they can be paired with atomistic simulations[69], or with crystal structures of alternative conformations and functional perturbations, as demonstrated here. In the future, functional perturbations such as the mutations that we used to study CypA, or binding of chemical ligands, could be paired with time-resolved structural data from temperature-jump experiments to provide insight into how disease alleles or drug molecules impinge on protein dynamics.

Supplementary Figures

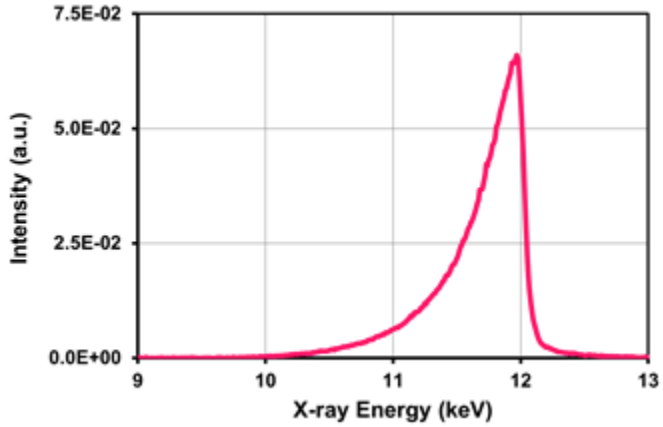

Figure 2.6. Pink beam energy spectrum. Typical X-ray energy spectrum of the pink beam (3% energy bandwidth) used for the reported SAXS/WAXS measurements.

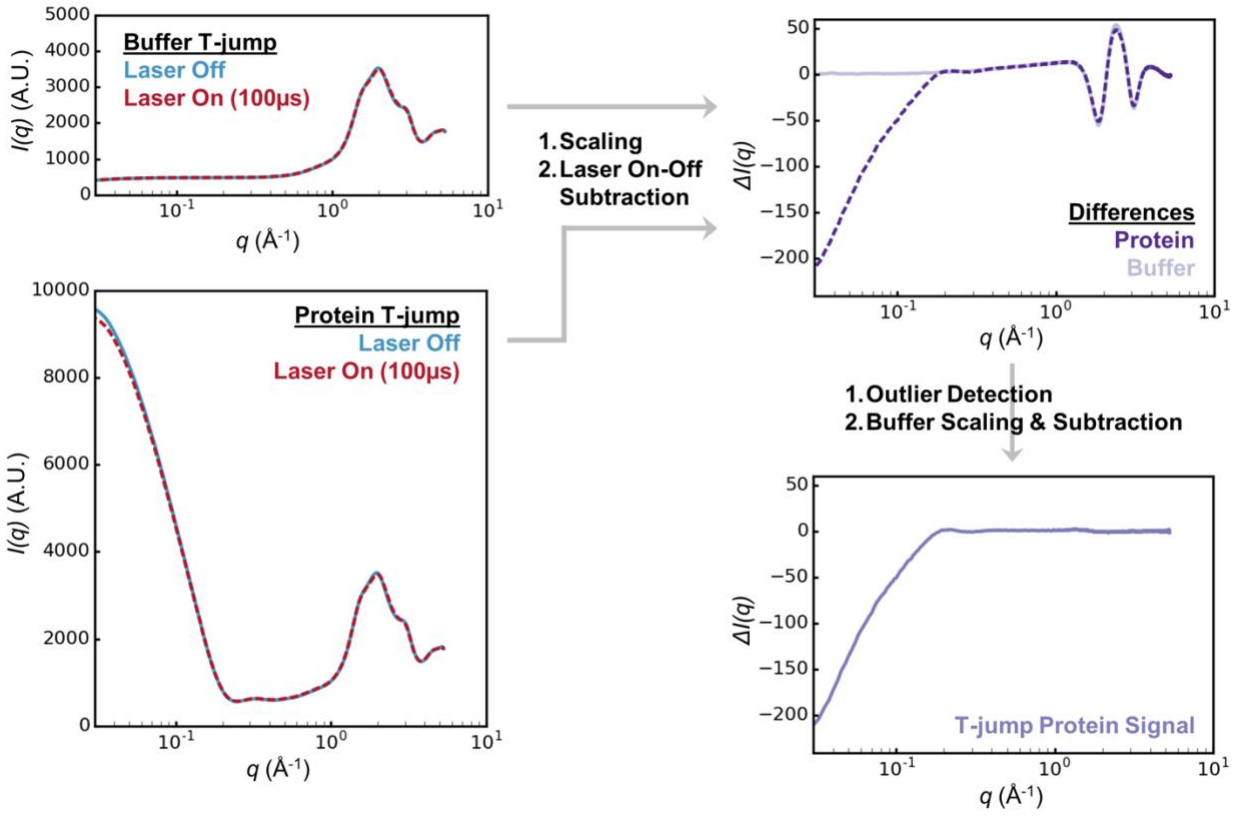

Figure 2.7. Reduction of T-jump SAXS/WAXS data. T-jump data processing involves a combination of scaling and subtraction operations that produce time-resolved difference scattering curves. For each “laser on-off” pair, the recorded scattering curves are scaled to one another and the “laser off” curve is then subtracted from the “laser on” curve. This procedure is

done independently for samples containing buffer only, and for protein samples. Next, the resulting difference curve for the buffer only sample is scaled to the difference curve obtained for the protein sample, and an additional buffer subtraction is performed to remove the thermal signal from the solvent. The result of this procedure is a difference scattering curve containing signal from the protein molecules only.

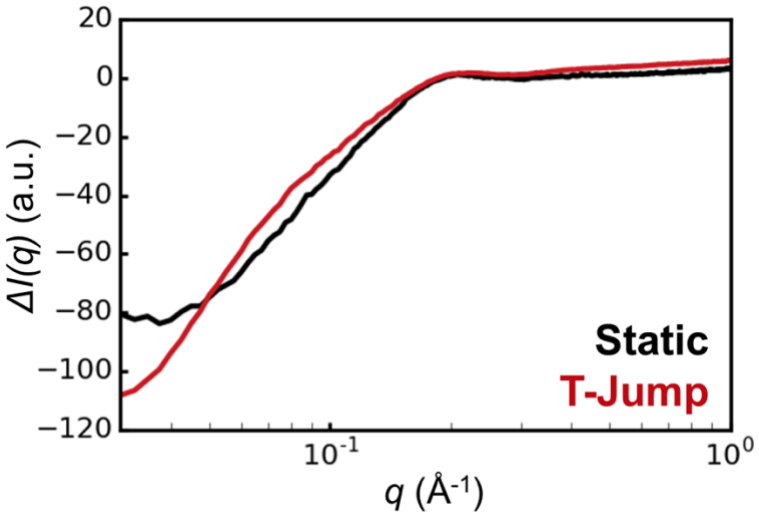

Figure 2.8. Comparison of static and time-resolved differences. Static scattering differences between CypA solutions at 13°C and 23°C (black curve), and time resolved differences (100μs-laser off) for a T-jump spanning a temperature range of approximately 15°C and 26°C (red curve) are shown, demonstrating that the time-resolved signal equilibrates to approximately what would be expected based on static measurements.

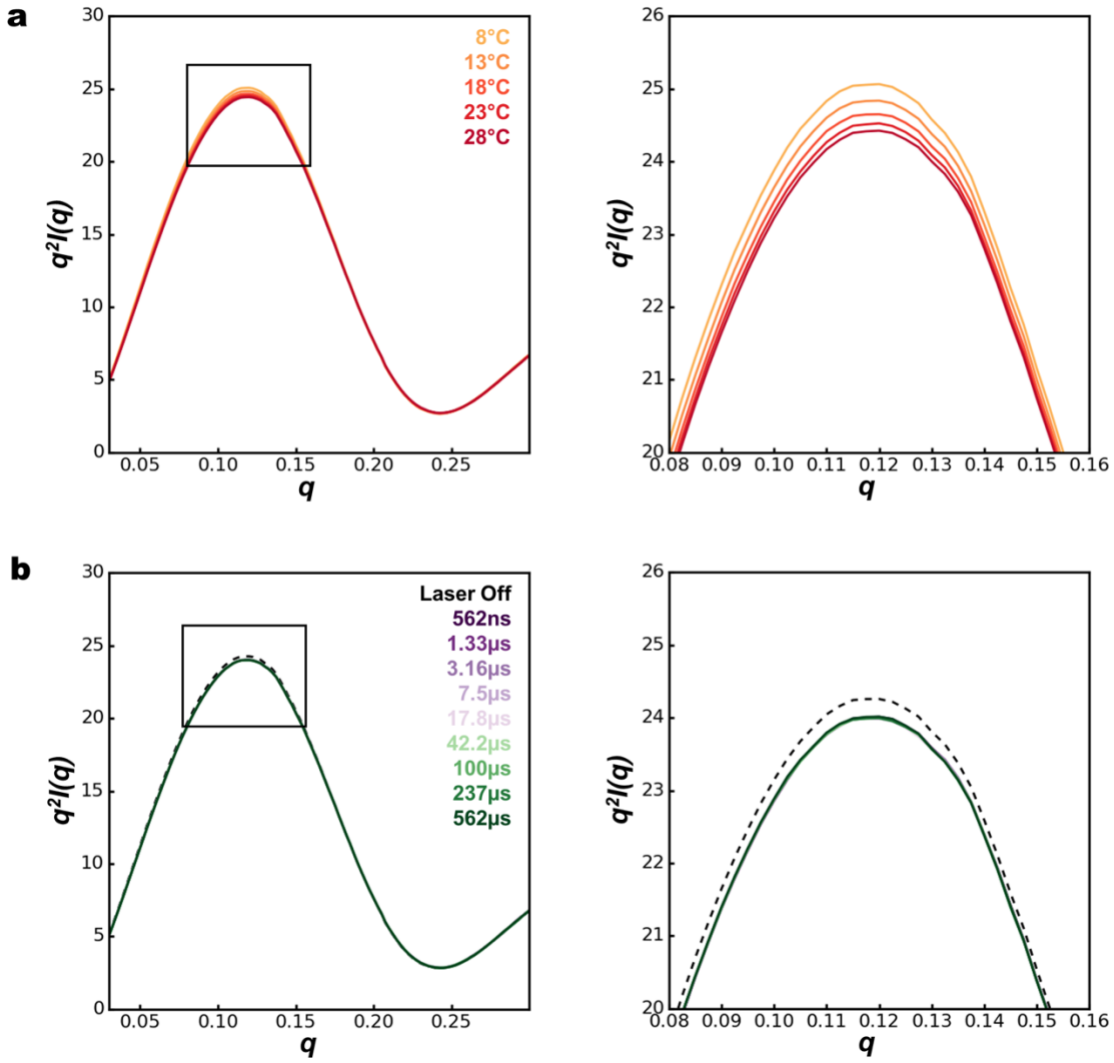

Figure 2.9. Kratky plots for CypA reveal a small thermal disorder effect without protein unfolding. **a)** Kratky plots calculated as a function of static temperature, from 8°C to 28°C. The right panel shows an expanded view of the boxed region. **b)** Kratky plots calculated as a function of time delay for time-resolved T-jump data (colored lines, T-jump from approximately 15°C to 26°C). Again, the right panel shows an expanded view of the boxed region. All time delays show a similar difference relative to the “laser off” state (black dashed line), indicating that the underlying structural change is faster than the measurement dead time of our experiment.

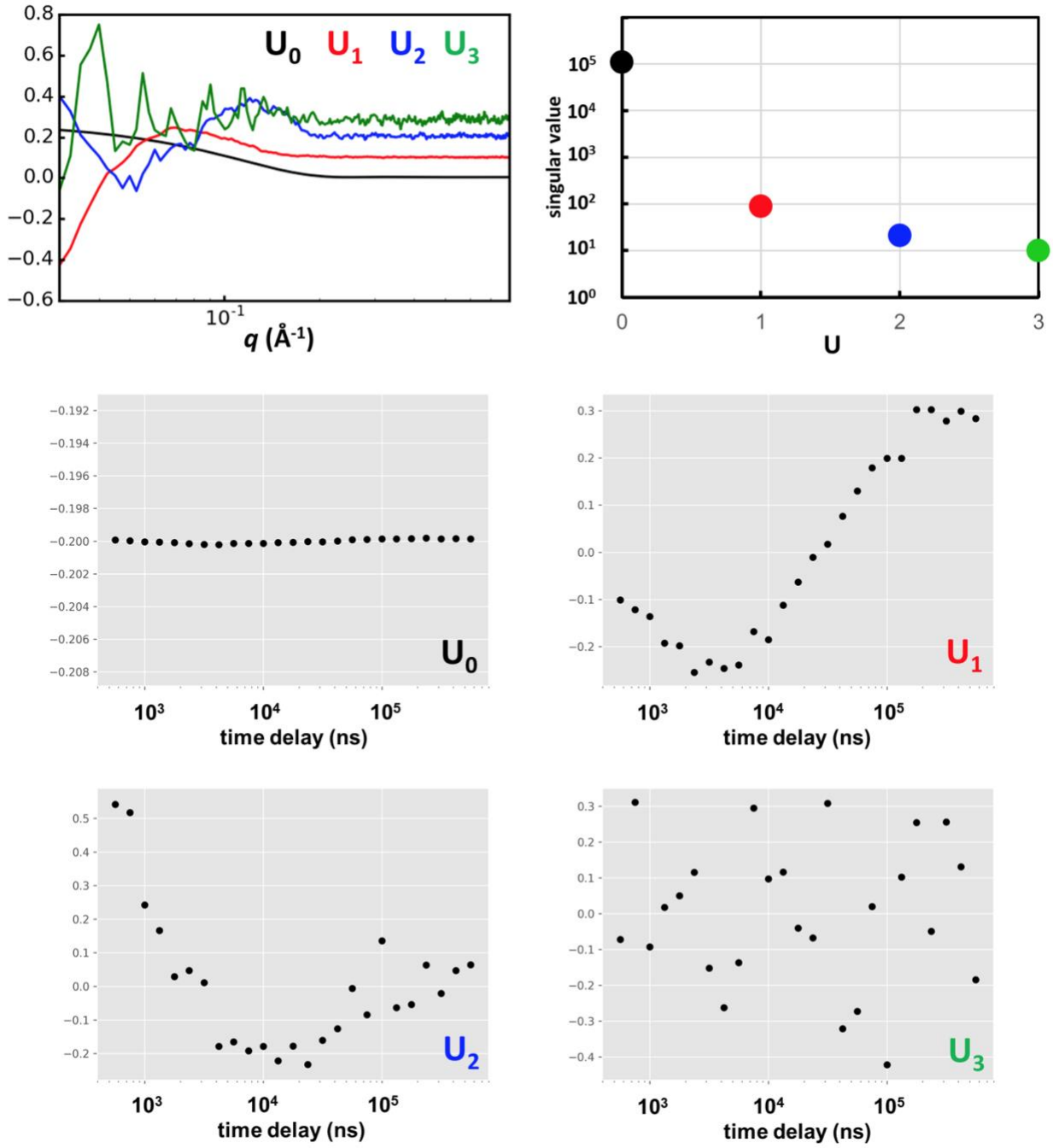

Figure 2.10. Kinetic analysis of X-ray scattering by singular value decomposition (SVD). We used SAXS curves representing the time-resolved scattering at each time delay spanning 562ns and 562μs to construct a matrix, which was analyzed by SVD. The top left panel shows the top four left singular vectors (Un), and the top right panel shows their corresponding singular values. The lower panels are constructed from the right singular vectors, and show the time-dependent contribution of each left singular vector to the total signal. We note that the primary time-resolved signal is contained in just one singular vector (U2), whose time-dependent behavior reflects that of the integration analysis described in the main text.

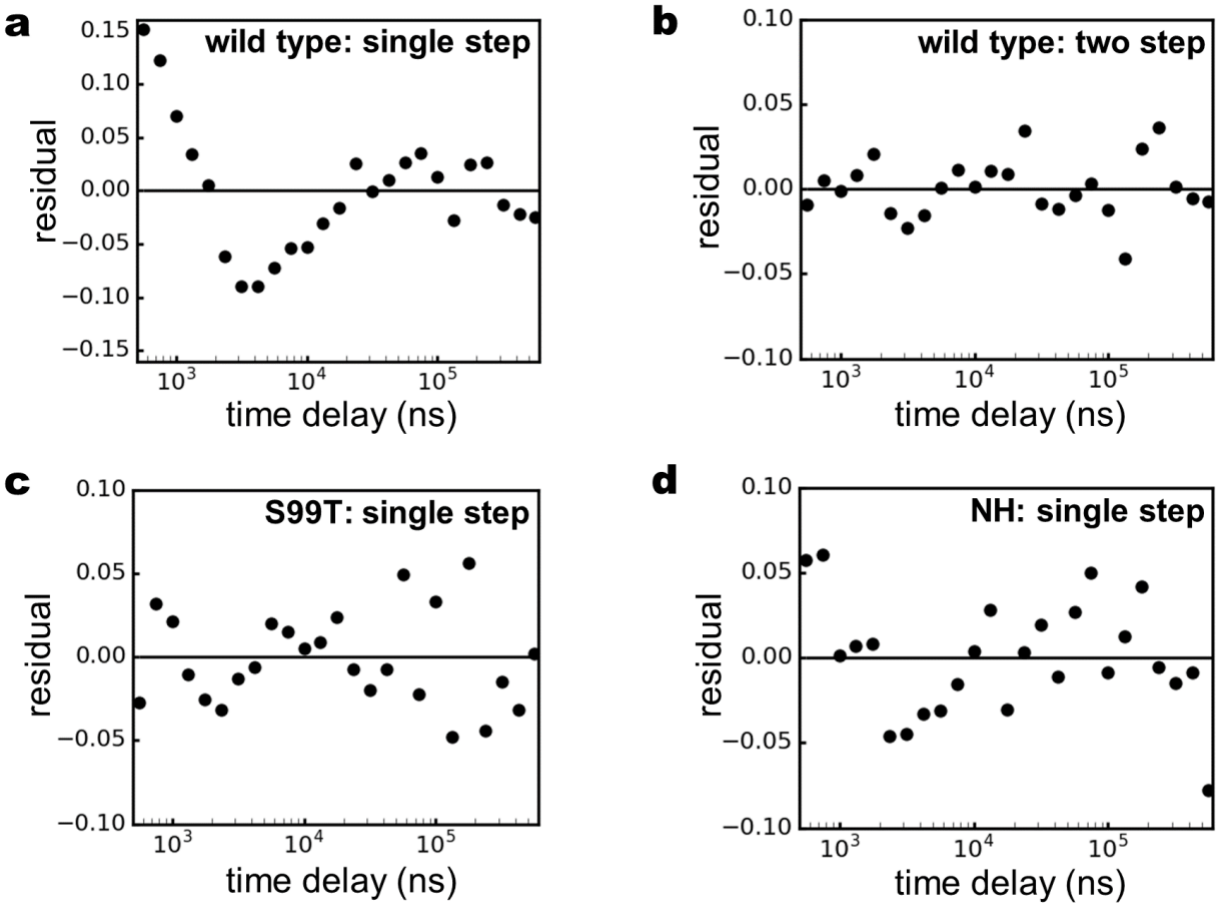

Figure 2.11. Residuals for kinetic fits of T-jump SAXS/WAXS data for CypA variants. Residuals are shown for the following variants and fitting procedures: **a)** Wild type, single-step relaxation. **b)** Wild type, two-step relaxation. **c)** S99T mutant, single-step relaxation. **d)** NH double mutant, single-step relaxation. When the wild type data are modeled using a single kinetic step **(a)**, the residuals show structure indicative of a second exponential process. In contrast, other fits show residuals with random fluctuations about zero **(b, c, d)**, indicating an appropriate number of steps were included in the kinetic models.

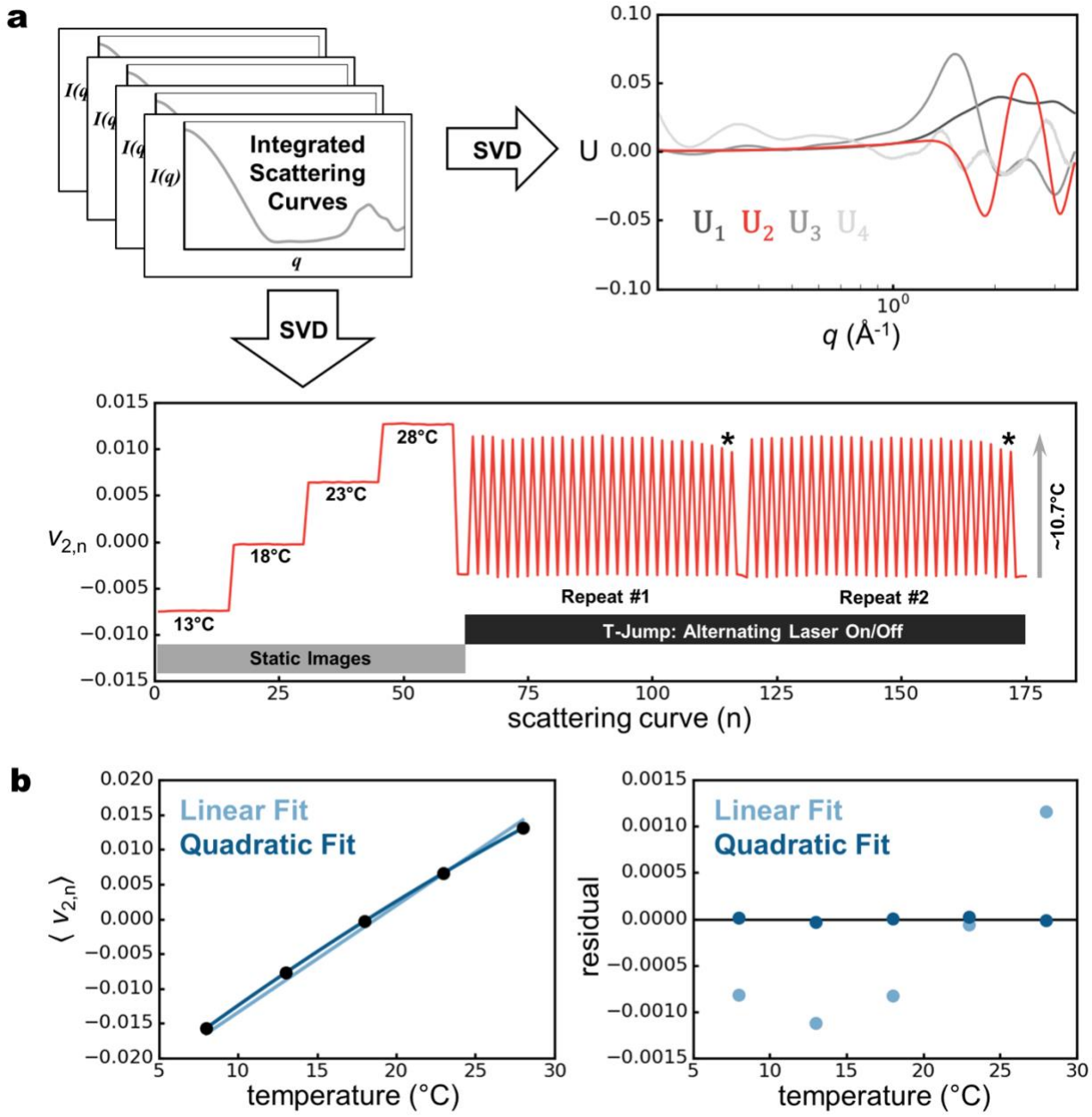

Figure 2.12. X-ray scattering from bulk water acts as a sensitive thermometer for T-jump experiments. a) Using singular value decomposition (SVD), we can identify a signal whose contribution to each scattering curve is strongly dependent on the temperature. The left singular vectors with the four highest singular values are shown, with the vector corresponding to the temperature-dependent signal (U2) colored red. The contribution of this vector (v2,n) to each of 175 scattering curves is also shown. This set of 175 scattering curves includes static measurements (no pump laser) at four different temperatures, followed by two repeats of time-resolved T-jump measurements. The T-jump data were collected as "laser on-off" pairs, and within a single repeat each successive on-off pair was collected with an increasing pump-probe time delay. Cooling is evident at longer pump-probe time delays (denoted by *). **b)** To calculate the magnitude of the laser-induced T-jump, we used the static data to determine the average value of v2,n as a function of temperature, and fit the data using both linear and quadratic

models. Based on the residuals for the two fits, we chose to use the resulting quadratic equation to determine the magnitude of the laser-induced T-jump using the values of v2,n calculated for the time-resolved scattering curves by SVD.

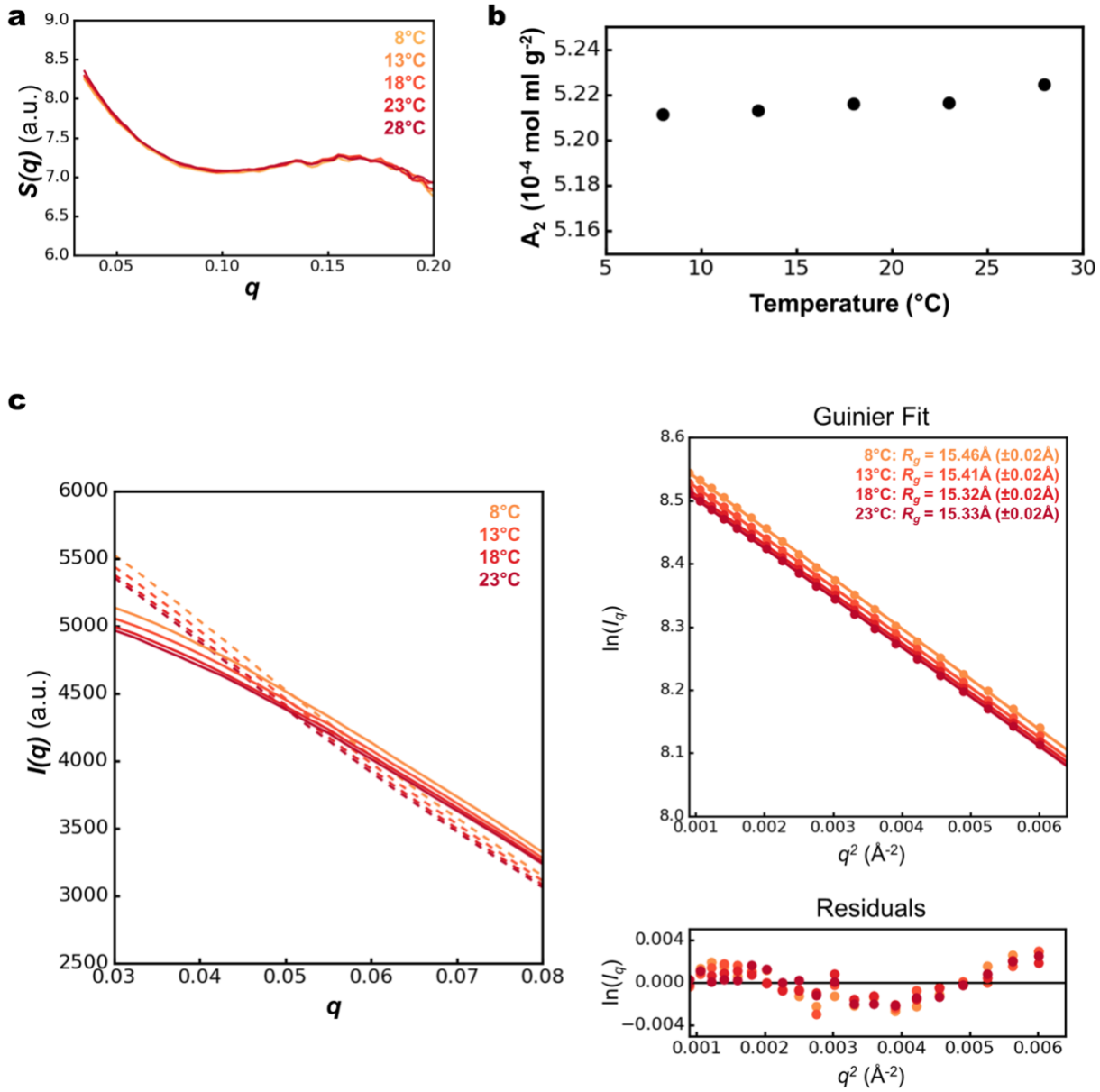

Figure 2.13. Intermolecular interactions are not temperature dependent for CypA solutions. Both structure factors (S(q) in **panel a)** and second virial coefficients (A2 in **panel b)** calculated for 50mg/mL CypA solutions (wild type) are unchanged over temperatures ranging from 8-28°C. Low angle scattering is affected by interparticle interactions (dashed lines in **panel c, left)**, however the effects can be removed by “infinite dilution extrapolation” (solid lines in **panel c, left)**. Guinier analysis (**panel c, right)** using the curves that are corrected for interparticle interactions shows that these effects can be successfully removed, and that the radius-of-gyration can be accurately determined at different temperatures.

Table 2.2. Calculated rates for the fast (k1) and slow (k2) relaxation processes measured from all T-jump experiments reported here. Note that kinetic analyses for the S99T and NH variants were performed at 25.5°C and 27.1°C respectively.

CypA Variant	Temperature (ºC)	Temperature (K)	k_1 (10^4 s^{-1})	k_2 (10^4 s^{-1})
WT	6.2 ± 0.2	279.2 ± 0.2	N/A	1.1± 0.3
WT	11.4 ± 0.2	284.4 ± 0.2	43 ± 7	1.2 ± 0.2
WT	15.9 ± 0.2	288.9 ± 0.2	49 ± 10	2.5 ± 0.3
WT	17.0 ± 0.1	290.0 ± 0.1	58 ± 11	1.6 ± 0.2
WT	21.5 ± 0.1	294.5 ± 0.1	57 ± 7	2.2 ± 0.2
WT	23.8 ± 0.2	296.8 ± 0.2	117 ± 33	2.2 ± 0.2
WT	25.5 ± 0.2	298.5 ± 0.2	53 ± 33	2.7 ± 0.9
WT	26.7 ± 0.1	299.7 ± 0.1	98 ± 12	2.5 ± 0.2
WT	29.9 ± 0.1	302.9 ± 0.1	116 ± 24	2.4 ± 0.2
S99T	25.5 ± 0.2	298.5 ± 0.2	34 ± 4	N/A
D66N/R69H	27.1 ± 0.1	300.1 ± 0.1	N/A	4.1 ± 0.6

Supplementary Data

Calibrating the Magnitude of the Temperature-Jump by Singular Value Decomposition

To characterize the temperature-dependent behavior of the solvent scattering, we performed static SAXS/WAXS measurements of our CypA samples as a function of temperature (equilibrium, no IR laser), in addition to our time-resolved measurements. We pooled these static, temperature-dependent SAXS/WAXS curves (azimuthally integrated I(q) v. q) with the time resolved SAXS/WAXS curves from our T-jump measurements, and performed singular value decomposition (SVD) on a matrix constructed from the set of pooled curves (**Figure 2.12a**). Specifically, each column of this matrix represents a scattering curve, with each row of the matrix corresponding to a q-bin and the entries in the matrix corresponding to measured scattering intensities. The SVD analysis, which was performed over the q=0.07-3.45Å-1 region

of the scattering curves, identified a signal (a left singular vector) whose prominent features were found in the q-region corresponding to the scattering of bulk water (q > 1.0) (**Figure 2.12a**). By extracting the entries in the corresponding row of the V matrix (containing the right singular vectors as columns), we could determine how this singular vector contributed to each scattering curve and demonstrate that its contribution was strongly temperature-dependent. Specifically, in the static (no T-jump) scattering curves the contribution of this singular vector increased with temperature, and for T-jump measurements the contribution of this vector to the observed scattering curves is perfectly correlated to the application of the pump laser pulse over sequential "laser on-off" pairs of X-ray measurements (**Figure 2.12a**), providing positive confirmation of a T-jump. Additionally, it is notable that the shape of the left singular vector used to monitor the temperature by SVD is nearly identical to that of the on-off difference signal at high-q.

The identification of a temperature-dependent singular vector provided a simple way to measure the magnitude of the laser-induced T-jump[9,25,27,47,48]. For each of the five static temperatures we explored, we calculated the average value of v2,n, the entry in the matrix V that describes the contribution of the temperature-dependent singular vector (U2) to the nth scattering curve, across 32 individual X-ray scattering curves. We then plotted the average v2,n vs. temperature and fit the data using both linear and quadratic models (**Figure 2.12b**). We examined the residuals for the two fits, determined the quadratic fit produced the most appropriate "standard curve" for estimation of the sample temperature from the SVD analysis, and used the resulting second-degree polynomial to estimate the temperature for each scattering curve in our series of time-resolved measurements. We compared the temperatures calculated for neighboring "laser on" and "laser" off scattering curves, and found the average T-jump produced by our IR heating pulse to be approximately 10.7°C. We observed that v2,n was consistent as a function of pump-probe time delay out to delay times of approximately 562μs, and decreased for longer time delays, implying that significant cooling of the sample occurred in

less than 1 millisecond. Consequently, we limited our subsequent analysis to time delays shorter than 562μs.

Additional details about the temperature calibration are provided in the Supplementary Methods section.

Characterization of Interparticle Interactions and Structure Factor Corrections

Because the main time-resolved signal change was confined to very low-q, we wanted to ensure our time-resolved signal was due to a change in the protein's form factor (infinite dilution), and not the structure factor of the protein solution. This process was challenging due to low signal-to-noise at high-q, as well as our inability to measure very low-angle scattering using the available instrumentation. However, static measurements allowed us to identify the contribution of interparticle scattering, demonstrate that this contribution is invariant to temperature over the range of our measurement conditions, and correct for its effect. To test whether changes in the radial distribution function originated from structural changes within the individual protein particles and their associated solvent and not from changes in the relative arrangement of the CypA molecules in solution, we performed static SAXS/WAXS measurements of CypA as a function of both temperature and CypA concentration. Using these data, we calculated the structure factor (S(q)) for a 50mg/mL CypA solution at multiple different temperatures (see Supplementary Methods), and determined there was no significant difference in the q=0.03-0.2 Å-1 region of the scattering curves (**Figure 2.13**), consistent with other work on similarly-sized protein molecules in solution[80]. Next, we plotted the second virial coefficient for CypA as a function of temperature, and noticed that this quantity shows only a very small temperature dependence that cannot account for the observed time-resolved differences. In contrast to our results for CypA, Bonneté, et al. performed similar calculations of second virial coefficients for lysozyme solutions at similar temperatures and buffer conditions, and calculated temperature-dependent changes that were 50-fold larger (or more) than what we determined for CypA[80]. This analysis also allowed us to correct for the effects of interparticle interactions

through a procedure which allows extrapolation of the observed concentration-dependent scattering to “infinite dilution” (see Supplementary Methods). In addition to the direct measurements of interparticle interactions provided by concentration-dependent scattering measurements, we also used Guinier analysis (linear fit of ln[I(q)] vs. q2, see Supplementary Methods) to assess whether the radial distribution function (structure factor) of CypA particles in solution changes significantly upon temperature perturbation. We analyzed scattering curves derived from both static temperature data and from time-resolved data, and observed that while Guinier fits of the T-jump measurements do have larger residuals than those for static data, the residuals do not change substantially as a function of either temperature (in static experiments) or time (in time-resolved experiments).

Rapid Loss of X-ray Scattering Contrast due to Thermal Expansion

We believe the largest contributor to the fast reduction in I(0) signal is thermal expansion, whose effect is twofold. First, thermal expansion of the system in general results in the expulsion of some scattering material (solvent containing CypA protein particles) from the X-ray beam path, which has a constant volume during the experiment. The volumetric thermal expansion coefficient of liquid water is approximately 2.1*10-4/K[81], and the sample is 95% water by mass, so the amount of material in the X-ray beam path decreases by approximately 0.2% for our T-jumps, which were approximately 11°C. Second, protein thermal expansion coefficients are estimated to be larger than those of liquid water, so heating of the sample results in a reduction of scattering density (loss of contrast) because the decrease in the average electron density of the protein particles is greater than that of the solvent. Using CypA crystal structures determined over a temperature range spanning 260-310K[43], we determined the thermal expansion coefficient of CypA to be approximately 4.1*10-4/K. The thermal expansion coefficient we calculated for CypA is similar to that of met-Myoglobin, which was reported to be 3.5*10-4/K by [82,83]. Using our calculated thermal expansion coefficient for CypA and its average electron density calculated from a crystal structure, along with the analogous

properties for bulk water[81], we estimate that the I(0) change due to thermal expansion is approximately 2.6% for our T-jump experiments; however, this approximation likely overestimates the effect of thermal expansion on the scattering density of CypA particles, because we calculated the thermal expansion coefficient of the protein using a linear fit of volume versus temperature over a broad temperature range. In contrast, the thermal expansion coefficient likely has some temperature dependence and our analysis may overestimate thermal expansion in the relevant temperature regime.

Supplementary Methods

Sample Preparation

CypA samples were prepared as described previously[43,44]. Briefly, the recombinant protein was expressed in E. coli BL21(DE3) cells and purified by liquid chromatography. Cells were lysed by sonication at pH=6.5, the lysate was clarified by high-speed centrifugation, and CypA was captured from the clarified lysate using a HiTrap-SP cation-exchange column (GE Healthcare). The protein was eluted using a sodium chloride gradient, and fractions containing CypA were pooled, and the pH was shifted to 7.5. The resulting solution was applied to a HiTrap-Q anion exchange column (GE Healthcare), and CypA was collected in the column flow-through. Finally, a polishing step was performed using a Superdex-75 gel filtration column (GE Healthcare). The protein was concentrated to 50mg/mL in buffer containing 20mM HEPES (4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid) buffer at pH=7.5, 50mM sodium chloride, and 0.5mM TCEP (trishydroxyethylphosphine). CypA mutants (S99T and NH) were prepared following the same protocol used for the wild type protein. We note that while 50mg/mL is a relatively high protein concentration for in vitro experiments, this is much lower than the typical intracellular protein concentration (approximately 300mg/mL), and that CypA has been demonstrated to undergo its functional motions even in the crystalline environment[43,44]. For all

X-ray measurements performed on buffer only without protein, the buffer was taken from the concentrator filtrate.

T-Jump SAXS/WAXS Data Collection and Processing

Time-resolved SAXS/WAXS measurements of CypA were performed on the BioCARS beamline at the Advanced Photon Source, while the storage ring was operating in hybrid mode. Temperature-jump data were acquired using the pump-probe method, as described recently by Cho. et al.[27]. Fast temperature-jump was performed on a CypA solution (50mg/mL) in a silica capillary using an Opolette 355 II (HE) optical parametric oscillator (OPOTEK), which produced a 7ns laser pulse with a peak wavelength of 1443nm. The pump laser energy was approximately 1mJ per pulse, and the beam was focused to an elliptical spot with dimensions of 400μm by 60μm (FWHM, gaussian beam profile), yielding a photon fluence of ~50mJ/mm2 at the sample, which heated a 50mg/mL CypA solution in a capillary. A suitably delayed X-ray pulse of 494ns duration (eight septuplets in APS hybrid mode) with a peak X-ray energy of 12keV and 3% energy bandwidth (pink beam, **Figure 2.6**), was used to probe the sample following the introduction of the T-jump, and the X-ray scattering was recorded using a Rayonix MX340-HS CCD detector. In our experiments, the temporal resolution is limited to approximately 500ns by the duration of the X-ray pulse, which is substantially longer than the duration of the IR pulse. To speed data acquisition, we utilized a sample holder and data collection scheme recently reported by Cho, et al,[27] which combined fast translation along the capillary axis with slow sample circulation via a peristaltic pump. The fast translation of the capillary allowed us to rapidly accumulate X-ray scattering from 41 pump-probe measurements on a single detector image by translating the capillary to a fresh position between each pump-probe pair. The slow circulation of the sample allowed us to replenish the protein solution and limit the extent of X-ray radiation damage by spreading the X-ray dose over a relatively large volume during long data collection runs. Data were collected as pairs of alternating "laser on" and "laser off" X-ray images. The pump-probe time delay was systematically increased with

each successive on/off pair of images. We measured pump-probe time delays spanning three logarithmic decades from 562ns to 1ms, at a time density of eight points per decade. A total of 50 replicate X-ray images were collected for each pump-probe time delay. It is important to note that time-resolved X-ray measurements referred to herein as “laser off,” were followed (10μs) by application of an IR pulse to the sample, as described by Cho, et al.[27], which prevented the introduction of a temperature offset created by incomplete cooling in between “laser on” and “laser off” measurements. A temperature controller integrated into the sample holder allowed us to initiate the T-jump from different starting temperatures, and also allowed us to collect static temperature data. Static temperature images were collected in a manner similar to the time-resolved images, but without application of the pump laser pulse. Data collection protocols were identical for protein and buffer samples.

After acquiring the data we applied polarization, geometry, and detector non-uniformity corrections to the 2D X-ray images. The scattering intensities (photons/pixel) were binned and averaged as a function of the scattering vector magnitude (q), yielding isotropic scattering curves (I(q) vs. q; $q = 4\pi \cdot \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength)[27]. Next, for each data collection run, we carried out outlier detection by performing singular value decomposition (SVD) on a matrix constructed from our integrated scattering curves. In this SVD, the left singular vector with the largest singular value represents the global average of all the scattering curves used to construct the input matrix. We analyzed the right singular vectors from the SVD to determine which images were irregular. Specifically, we calculated the mean value of v1,n, the entry in the matrix V that describes the contribution of the right singular vector with the largest singular value (U1) to the nth scattering curve, across all the input X-ray scattering curves. Then, if the value of v1,n for any specific scattering was more than 2.5 standard deviations above or below the mean, that image was discarded. Our outlier detection procedure is implemented in a Python script called “SVD_Quarantine.py.” By inspecting the results of the SVD, we decided to remove the first 5 repeats from each data set,

as well as some additional outliers that failed our statistical test. The same averaging and outlier detection method was used for both static and time-resolved measurements.

Scaling of X-ray Scattering Curves

All scaling of X-ray scattering curves was performed using an algebraic (least-squares) procedure. To determine the scale factor, A, which can be applied to a scattering curve I(q)a in order to scale it to a second scattering curve I(q)b, we used the following equation:

$$A = \frac{\Sigma_q I(q)_a I(q)_b}{\Sigma_q I(q)_a^2} \quad \textbf{(Eq. 2.2)}$$

Although we used the equation above for scaling throughout our analysis, the q-range to which it was applied varied depending on the context, and details are provided below.

X-ray Thermometry

Following the initial data processing steps described above, we used singular value decomposition (SVD) to determine the magnitude of the T-jump introduced by the IR laser pulse[9,25,27,47,48]. We pooled static, temperature-dependent SAXS/WAXS curves (azimuthally integrated I(q) v. q) with the time-resolved SAXS/WAXS curves from time-resolved measurements, scaled them to a common reference over the q=0.025-4.28 Å-1 region, and performed SVD on a matrix built from these scaled curves. In this matrix, each column represents a single scattering curve, with the rows of the matrix corresponding to q-bins and the entries in the matrix consisting of azimuthally averaged scattering intensities. The SVD analysis was performed using only the q=0.07-3.45 Å-1 region of the scattering curves. As described in the Results section, the SVD identified a left singular vector whose contribution to the overall scattering signal was highly temperature dependent. This was the left singular vector with the second largest singular value (U2). For each of the five temperatures used for static data collection, we calculated the average value of v2,n, which is the entry in the matrix V describing the contribution of the temperature-dependent singular vector (U2) to the nth scattering curve. We then plotted the average v2,n vs. temperature and ultimately fit this data using a quadratic

model. Finally, we used the resulting second-degree polynomial and the values v2,n for each time-resolved scattering curve to estimate the temperature for each T-jump measurement. By comparison of neighboring “laser on” and “laser off” scattering curves, we determined that the average T-jump was 10.7°C. Our thermometry procedure is implemented in a Python script called “thermometry_timepoints.py.”

Data reduction: On-Off Subtraction, Repeat Averaging, and Buffer Subtraction

We implemented a data reduction procedure that operated on the integrated scattering curves generated using our data collection protocol and produced several outputs that were subsequently used for our kinetic and structural analyses. This procedure, implemented in a Python script called “reduce_data.py,” took advantage of paired “laser on” and “laser off” measurements, redundant measurements of each pump-probe time-delay, and parallel T-jump experiments for samples containing protein and samples consisting of buffer only. The input for this script was essentially two data sets. The first, was a series of time-resolved scattering curves measured from a sample containing protein and consisting of paired “laser on/off” measurements with multiple replicate measurements of each pump-probe time delay (see above). The second was a similar data set, only collected from a sample containing buffer only and no protein. All of the input scattering curves were scaled to a common reference over the q=0.025-4.28Å-1 range, and “laser off” curves were subtracted from their associated “laser on” curves to create a difference scattering curve (ΔI(q)) for each “laser on/off” pair. Next, all replicate difference curves (i.e. same sample and time delay) were grouped together, an iterative chi-squared test was performed (using a cutoff of $\chi 2=1.5$), and the average difference curve was calculated for each pump-probe time delay in the series. For each time delay, the difference signal for the buffer only sample was scaled to the difference signal for the sample containing protein over the q=1.5-3.6 Å-1 range, and then the buffer signal was subtracted from the protein signal to isolate the difference signal due only to the protein. Additionally, this script took all of the “laser off” scattering curves, performed an iterative chi-squared test (cutoff of

χ2=1.5), and calculated their average. As was done for the difference curves, the average "laser off" scattering curve for buffer only was subtracted from the average "laser off" scattering curve for the protein sample after an additional scaling step (q=1.5-3.6 Å-1 range). The output of this data reduction procedure was a single scattering curve (I(q) vs. q) for the "laser off" state, and a difference scattering curve (ΔI(q) vs. q) for each pump-probe time delay. All output data were corrected for the contribution of the buffer, and errors were propagated from the initial measurement standard deviations.

Kinetic Analysis

The averaged difference curves produced by our data reduction procedure were used for kinetic analysis of the time-resolved signal changes, which was implemented in a Python script called "difference_dat_kinetics.py." For each time delay, this script integrated the area under the difference curve over the q=0.03-0.05 Å-1 range, then fit the resulting data (integrated area vs. time) to calculate relaxation rates using non-linear least-squares curve fitting. We used the following equations, for single-step kinetic fits:

$$A\left(1 - e^{-k_1 t}\right) + B \qquad \textbf{(Eq. 2.3)}$$

And for two-step kinetic fits:

$$A\left(1 - e^{-k_1 t}\right) + B\left(1 - e^{-k_2 t}\right) + C \qquad \textbf{(Eq. 2.4)}$$

The output of this analysis was a relaxation rate, or two rates, with standard deviations calculated using a bootstrapping method[51]. The use of a bootstrap procedure was deemed appropriate, because standard deviations calculated for the integrated area under the difference scattering curves are likely to overestimate the true error, as they are the result of propagating measurement standard deviations through radial integration, scaling, on-off subtraction, averaging, buffer subtraction, and difference curve integration, with all experimental errors assumed to be random. In contrast, some experimental error is likely systematic, and would instead be removed, rather than propagated, by the subtractive operations employed during

data processing. In cases where we performed T-jumps at multiple temperatures, we used the observed rates and their standard deviations to perform an Eyring analysis by fitting (Eq. 2.1) (above) using a least-squares method to determine the enthalpy and entropy of activation, and their standard deviations (using the covariance matrix). We implemented the Eyring analysis in a Python script called “eyring_fit.py.”

Creation of High-Quality Time-Resolved Scattering Curves for Structural Analysis

To produce high-quality scattering curves that could be used for real space interpretation of the time-resolved X-ray scattering, we took the following steps. We used all of the “laser off” scattering curves from our on-off paired time-resolved measurements to create a single average curve. Then, for each of the time-delays reported, we added the average on-off difference (see above) to this average “laser off” scattering curve:

$$I(q)_t = < I(q)_{off} > + < \Delta I(q)_t > \qquad \textbf{(Eq. 2.5)}$$

Calculating the time-resolved scattering curves in this manner allows us to take advantage of the data collection structure, specifically the collection of “laser on and off” pairs, and many replicate measurements to remove some effects of systematic error that can accumulate during the course of long experiments and can mask small time-resolved differences in X-ray scattering. Next, we utilized static scattering measurements, as a function of both concentration and temperature, to characterize the effect of intermolecular interactions on the observed X-ray scattering and to calculate structure factors (S(q)) for our 50mg/mL CypA solutions at temperatures spanning a range relevant to our T-jump experiments. We calculated structure factors (and second virial coefficients) following the methods described by Bonneté, et al.[80]. The scattering curves derived from summing the average “laser off” signal and the time-resolved differences were then divided by the calculated structure factors to correct for intermolecular interactions and extrapolate our measurements to infinite dilution. Because we discovered that the effect of intermolecular interactions was not temperature dependent, we did not need to model the time-dependence of structure factors for our protein solutions following the T-jump,

and the structure factor calculated at 13°C was used for the infinite dilution extrapolation. The calculation of structure factors and the creation of the high-quality, corrected I(q) curves were implemented in a pair of Python scripts called “structure_factor_calc.py” and “structure_factor_correction_TR.py,” respectively (a script called “structure_factor_correction_static.py,” was used in the latter step for static data sets).

Guinier Analysis and Calculation of Rg

In order to calculate radii-of-gyration (Rg) and to extrapolate the value of I(0) from scattering curves, we used the linear Guinier approximation:

$$ln[I(q)] = ln[I(0)] - \frac{R_g^2}{3} \cdot q^2 \qquad \textbf{(Eq. 2.6)}$$

Guinier analysis was performed over the q-region spanning 0.03-0.08Å-1. We note that scattering curves were not placed on an absolute scale, however, this is not a requirement for Guinier analysis. The calculations were implemented in a Python script called “rg_and_i0.py.”

Data availability

Scattering data are deposited at NIH Figshare (https://doi.org/10.35092/yhjc.9177143). Additional information and files are available from the corresponding author upon reasonable request.

Code availability

All Python scripts used for analysis of integrated X-ray scattering curves are publicly available from GitHub (https://github.com/fraser-lab/solution_scattering). A code release checkpoint containing the exact scripts used in this work is available via Zenodo (https://doi.org/10.5281/zenodo.3355707).

References

1. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
2. Bedem, H. van den & Fraser, J. S. Integrative, dynamic structural biology at atomic resolution—it's about time. *Nat. Methods* **12**, 307–318 (2015).
3. Bottaro, S. & Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **361**, 355–360 (2018).
4. Schmidt, M. Time-Resolved Macromolecular Crystallography at Modern X-Ray Sources. in *Protein Crystallography: Methods and Protocols* (eds. Wlodawer, A., Dauter, Z. & Jaskolski, M.) 273–294 (Springer, 2017). doi:10.1007/978-1-4939-7000-1_11.
5. Neutze, R. & Moffat, K. Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges. *Curr. Opin. Struct. Biol.* **22**, 651–659 (2012).
6. Schotte, F. *et al.* Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. *Proc. Natl. Acad. Sci.* **109**, 19256–19261 (2012).
7. Cho, H. S. *et al.* Picosecond Photobiology: Watching a Signaling Protein Function in Real Time via Time-Resolved Small- and Wide-Angle X-ray Scattering. *J. Am. Chem. Soc.* **138**, 8815–8823 (2016).
8. Schlichting, I. & Miao, J. Emerging opportunities in structural biology with X-ray free-electron lasers. *Curr. Opin. Struct. Biol.* **22**, 613–626 (2012).
9. Arnlund, D. *et al.* Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nat. Methods* **11**, 923–926 (2014).
10. Berntsson, O. *et al.* Sequential conformational transitions and α-helical supercoiling regulate a sensor histidine kinase. *Nat. Commun.* **8**, 1–8 (2017).
11. Takala, H. *et al.* Signal amplification and transduction in phytochrome photosensors. *Nature* **509**, 245–248 (2014).
12. Brinkmann, L. U. L. & Hub, J. S. Ultrafast anisotropic protein quake propagation after CO

photodissociation in myoglobin. *Proc. Natl. Acad. Sci.* **113**, 10565–10570 (2016).

13. Barends, T. R. M. *et al.* Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350**, 445–450 (2015).
14. Coquelle, N. *et al.* Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* **10**, 31–37 (2018).
15. Pande, K. *et al.* Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725–729 (2016).
16. Kern, J. *et al.* Structures of the intermediates of Kok's photosynthetic water oxidation clock. *Nature* **563**, 421–425 (2018).
17. Nogly, P. *et al.* Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* **361**, (2018).
18. Malmerberg, E. *et al.* Time-Resolved WAXS Reveals Accelerated Conformational Changes in Iodoretinal-Substituted Proteorhodopsin. *Biophys. J.* **101**, 1345–1353 (2011).
19. Hekstra, D. R. *et al.* Electric-field-stimulated protein mechanics. *Nature* **540**, 400–405 (2016).
20. Schlichting, I. *et al.* Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature* **345**, 309–315 (1990).
21. Stoddard, B. L., Cohen, B. E., Brubaker, M., Mesecar, A. D. & Koshland, D. E. Millisecond Laue structures of an enzyme–product complex using photocaged substrate analogs. *Nat. Struct. Biol.* **5**, 891–897 (1998).
22. Josts, I. *et al.* Photocage-initiated time-resolved solution X-ray scattering investigation of protein dimerization. *IUCrJ* **5**, 667–672 (2018).
23. Rimmerman, D. *et al.* Revealing Fast Structural Dynamics in pH-Responsive Peptides with Time-Resolved X-ray Scattering. *J. Phys. Chem. B* **123**, 2016–2021 (2019).
24. Olmos, J. L. *et al.* Enzyme intermediates captured "on the fly" by mix-and-inject serial

crystallography. *BMC Biol.* **16**, 59 (2018).

25. Rimmerman, D. *et al.* Direct Observation of Insulin Association Dynamics with Time-Resolved X-ray Scattering. *J. Phys. Chem. Lett.* **8**, 4413–4418 (2017).
26. Rimmerman, D. *et al.* Insulin hexamer dissociation dynamics revealed by photoinduced T-jumps and time-resolved X-ray solution scattering. *Photochem. Photobiol. Sci.* **17**, 874–882 (2018).
27. Cho, H. S. *et al.* Dynamics of Quaternary Structure Transitions in R-State Carbonmonoxyhemoglobin Unveiled in Time-Resolved X-ray Scattering Patterns Following a Temperature Jump. *J. Phys. Chem. B* **122**, 11488–11496 (2018).
28. Frauenfelder, H., Fenimore, P. W. & Young, R. D. Protein dynamics and function: Insights from the energy landscape and solvent slaving. *IUBMB Life* **59**, 506–512 (2007).
29. Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Parak, F. G. Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc. Natl. Acad. Sci.* **99**, 16047–16051 (2002).
30. Wang, J. & El-Sayed, M. A. Temperature Jump-Induced Secondary Structural Change of the Membrane Protein Bacteriorhodopsin in the Premelting Temperature Region: A Nanosecond Time-Resolved Fourier Transform Infrared Study. *Biophys. J.* **76**, 2777–2783 (1999).
31. Wang, T., Lau, W. L., DeGrado, W. F. & Gai, F. T-Jump Infrared Study of the Folding Mechanism of Coiled-Coil GCN4-p1. *Biophys. J.* **89**, 4180–4187 (2005).
32. Akasaka, K., Naito, A. & Nakatani, H. Temperature-jump NMR study of protein folding: Ribonuclease A at low pH. *J. Biomol. NMR* **1**, 65–70 (1991).
33. Gillespie, B. *et al.* NMR and Temperature-jump Measurements of de Novo Designed Proteins Demonstrate Rapid Folding in the Absence of Explicit Selection for Kinetics. *J. Mol. Biol.* **330**, 813–819 (2003).
34. Yamasaki, K. *et al.* Real-Time NMR Monitoring of Protein-Folding Kinetics by a Recycle

Flow System for Temperature Jump. *Anal. Chem.* **85**, 9439–9443 (2013).

35. Meadows, C. W., Balakrishnan, G., Kier, B. L., Spiro, T. G. & Klinman, J. P. Temperature-Jump Fluorescence Provides Evidence for Fully Reversible Microsecond Dynamics in a Thermophilic Alcohol Dehydrogenase. *J. Am. Chem. Soc.* **137**, 10060–10063 (2015).
36. Vaughn, M. B., Zhang, J., Spiro, T. G., Dyer, R. B. & Klinman, J. P. Activity-Related Microsecond Dynamics Revealed by Temperature-Jump Förster Resonance Energy Transfer Measurements on Thermophilic Alcohol Dehydrogenase. *J. Am. Chem. Soc.* **140**, 900–903 (2018).
37. Hori, T. *et al.* The initial step of the thermal unfolding of 3-isopropylmalate dehydrogenase detected by the temperature-jump Laue method. *Protein Eng. Des. Sel.* **13**, 527–533 (2000).
38. Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121 (2005).
39. Caines, M. E. C. *et al.* Diverse HIV viruses are targeted by a conformationally dynamic antiviral. *Nat. Struct. Mol. Biol.* **19**, 411–416 (2012).
40. Price, A. J. *et al.* Active site remodeling switches HIV specificity of antiretroviral TRIMCyp. *Nat. Struct. Mol. Biol.* **16**, 1036–1042 (2009).
41. Virgen, C. A., Kratovac, Z., Bieniasz, P. D. & Hatziioannou, T. Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3563–3568 (2008).
42. Wilson, S. J. *et al.* Independent evolution of an antiviral TRIMCyp in rhesus macaques. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3557–3562 (2008).
43. Keedy, D. A. *et al.* Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife* **4**, e07574 (2015).
44. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).

45. Otten, R. *et al.* Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat. Commun.* **9**, 1–11 (2018).
46. Clark, G. N. I., Hura, G. L., Teixeira, J., Soper, A. K. & Head-Gordon, T. Small-angle scattering and the structure of ambient liquid water. *Proc. Natl. Acad. Sci.* **107**, 14003–14007 (2010).
47. Skov Kjær, K. *et al.* Introducing a standard method for experimental determination of the solvent response in laser pump, X-ray probe time-resolved wide-angle X-ray scattering experiments on systems in solution. *Phys. Chem. Chem. Phys.* **15**, 15003–15016 (2013).
48. Cammarata, M. *et al.* Impulsive solvent heating probed by picosecond x-ray diffraction. *J. Chem. Phys.* **124**, 124504 (2006).
49. Gruebele, M., Sabelko, J., Ballew, R. & Ervin, J. Laser Temperature Jump Induced Protein Refolding. *Acc. Chem. Res.* **31**, 699–707 (1998).
50. Levantino, M. *et al.* Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. *Nat. Commun.* **6**, 1–6 (2015).
51. DiCiccio, T. J. & Efron, B. Bootstrap Confidence Intervals. *Stat. Sci.* **11**, 189–212 (1996).
52. Smith, C. A. *et al.* Population Shuffling of Protein Conformations. *Angew. Chem. Int. Ed.* **54**, 207–210 (2015).
53. Wapeesittipan, P., Mey, A. S. J. S., Walkinshaw, M. D. & Michel, J. Allosteric effects in catalytic impaired variants of the enzyme cyclophilin A may be explained by changes in nano-microsecond time scale motions. *bioRxiv* 224329 (2018) doi:10.1101/224329.
54. Chi, C. N. *et al.* A Structural Ensemble for the Enzyme Cyclophilin Reveals an Orchestrated Mode of Action at Atomic Resolution. *Angew. Chem. Int. Ed.* **54**, 11657–11661 (2015).
55. Kim, J. G. *et al.* Cooperative protein structural dynamics of homodimeric hemoglobin linked to water cluster at subunit interface revealed by time-resolved X-ray solution scattering. *Struct. Dyn.* **3**, 023610 (2016).
56. Colombo, M. F., Rau, D. C. & Parsegian, V. A. Protein solvation in allosteric regulation: a

water effect on hemoglobin. *Science* **256**, 655–659 (1992).

57. Salvay, A. G., Grigera, J. R. & Colombo, M. F. The Role of Hydration on the Mechanism of Allosteric Regulation: In Situ Measurements of the Oxygen-Linked Kinetics of Water Binding to Hemoglobin. *Biophys. J.* **84**, 564–570 (2003).
58. Royer, W. E., Pardanani, A., Gibson, Q. H., Peterson, E. S. & Friedman, J. M. Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proc. Natl. Acad. Sci.* **93**, 14526–14531 (1996).
59. Fenwick, R. B., Oyen, D., Dyson, H. J. & Wright, P. E. Slow Dynamics of Tryptophan–Water Networks in Proteins. *J. Am. Chem. Soc.* **140**, 675–682 (2018).
60. Grossman, M. *et al.* Correlated structural kinetics and retarded solvent dynamics at the metalloprotease active site. *Nat. Struct. Mol. Biol.* **18**, 1102–1108 (2011).
61. Decaneto, E. *et al.* Solvent water interactions within the active site of the membrane type I matrix metalloproteinase. *Phys. Chem. Chem. Phys.* **19**, 30316–30331 (2017).
62. Leidner, F., Kurt Yilmaz, N., Paulsen, J., Muller, Y. A. & Schiffer, C. A. Hydration Structure and Dynamics of Inhibitor-Bound HIV-1 Protease. *J. Chem. Theory Comput.* **14**, 2784–2796 (2018).
63. Guha, S. *et al.* Slow Solvation Dynamics at the Active Site of an Enzyme: Implications for Catalysis. *Biochemistry* **44**, 8940–8947 (2005).
64. Dahanayake, J. N. & Mitchell-Koch, K. R. Entropy connects water structure and dynamics in protein hydration layer. *Phys. Chem. Chem. Phys.* **20**, 14765–14777 (2018).
65. Wand, A. J. & Sharp, K. A. Measuring Entropy in Molecular Recognition by Proteins. *Annu. Rev. Biophys.* **47**, 41–61 (2018).
66. Caro, J. A. *et al.* Entropy in molecular recognition by proteins. *Proc. Natl. Acad. Sci.* **114**, 6563–6568 (2017).
67. Gavrilov, Y., Leuchter, J. D. & Levy, Y. On the coupling between the dynamics of protein and water. *Phys. Chem. Chem. Phys.* **19**, 8243–8257 (2017).

68. Conti Nibali, V., D'Angelo, G., Paciaroni, A., Tobias, D. J. & Tarek, M. On the Coupling between the Collective Dynamics of Proteins and Their Hydration Water. *J. Phys. Chem. Lett.* **5**, 1181–1186 (2014).
69. Hub, J. S. Interpreting solution X-ray scattering data using molecular simulations. *Curr. Opin. Struct. Biol.* **49**, 18–26 (2018).
70. Svergun, D. I. *et al.* Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci.* **95**, 2267–2272 (1998).
71. Virtanen, J. J., Makowski, L., Sosnick, T. R. & Freed, K. F. Modeling the Hydration Layer around Proteins: Applications to Small- and Wide-Angle X-Ray Scattering. *Biophys. J.* **101**, 2061–2069 (2011).
72. Henriques, J., Arleth, L., Lindorff-Larsen, K. & Skepö, M. On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J. Mol. Biol.* **430**, 2521–2539 (2018).
73. Dellus-Gur, E. *et al.* Negative Epistasis and Evolvability in TEM-1 β-Lactamase—The Thin Line between an Enzyme's Conformational Freedom and Disorder. *J. Mol. Biol.* **427**, 2396–2409 (2015).
74. Biel, J. T., Thompson, M. C., Cunningham, C. N., Corn, J. E. & Fraser, J. S. Flexibility and Design: Conformational Heterogeneity along the Evolutionary Trajectory of a Redesigned Ubiquitin. *Structure* **25**, 739-749.e3 (2017).
75. Fischer, M., Coleman, R. G., Fraser, J. S. & Shoichet, B. K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat. Chem.* **6**, 575–583 (2014).
76. Fischer, M., Shoichet, B. K. & Fraser, J. S. One Crystal, Two Temperatures: Cryocooling Penalties Alter Ligand Binding to Transient Protein Sites. *ChemBioChem* **16**, 1560–1564 (2015).
77. Keedy, D. A. *et al.* An expanded allosteric network in PTP1B by multitemperature

crystallography, fragment screening, and covalent tethering. *eLife* **7**, e36307 (2018).

78. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci.* **108**, 16247–16252 (2011).
79. Bedem, H. van den, Bhabha, G., Yang, K., Wright, P. E. & Fraser, J. S. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat. Methods* **10**, 896–902 (2013).
80. Bonneté, F., Finet, S. & Tardieu, A. Second virial coefficient: variations with lysozyme crystallization conditions. *J. Cryst. Growth* **196**, 403–414 (1999).
81. Irvine, T. F. & Duignan, M. R. Isobaric thermal expansion coefficients for water over large temperature and pressure ranges. *Int. Commun. Heat Mass Transf.* **12**, 465–478 (1985).
82. Frauenfelder, H. *et al.* Thermal expansion of a protein. *Biochemistry* **26**, 254–261 (1987).
83. Hiebl, M. & Maksymiw, R. Anomalous temperature dependence of the thermal expansion of proteins. *Biopolymers* **31**, 161–167 (1991).

Acknowledgments

We thank R. Ranganathan, J. Holton, G. Hura and D. Elnatan for helpful discussions, and the staff at the BioCARS beamline at the Advanced Photon Source (I. Kosheleva, R. Henning, A. DiChiara and V. Srajer) for assistance. This work was supported by the NSF (STC-1231306), the NIH (GM123159 and GM124149), a Packard Fellowship from the David and Lucile Packard Foundation, the UC Office of the President Laboratory Fees Research Program LFR-17-476732 (to J.S.F.), the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases (to P.A.) and a Ruth L. Kirschstein National Research Service Award (F32 HL129989 to M.C.T.). Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract number DE-AC02-06CH11357. Use of the BioCARS Sector 14 was also supported by the NIH National Institute of General Medical Sciences (grant R24GM111072). The time-resolved setup at Sector 14 was funded in part through a collaboration with P.A. (NIH/NIDDK).

Contributions

M.C.T., P.A. and J.S.F. conceived and designed the experiments. M.C.T., B.A.B., A.M.W., H.S.C., F.S., D.M.C.S. and P.A. performed the experiments. M.C.T., B.A.B. and A.M.W. analyzed the data. M.C.T., B.A.B., A.M.W., H.S.C., F.S. and P.A. contributed materials/analysis tools. M.C.T. and J.S.F. wrote the paper. All authors discussed the results and commented on the manuscript.

Chapter III

Expansion of X-ray temperature-jump detection methods for time-resolved crystallography

Contributing Authors

Alexander M. Wolff, Michael C. Thompson, Benjamin A. Barad, Aaron S. Brewster, Iris D. Young, Eriko Nango, Michihiro Sugahara, Takanori Nakane, Kazutaka Ito, Tomoya Hino, Dohyun Im, Minoru Kubo, Takashi Nomura, Shigeki Owada, Tomoyuki Tanaka, Rie Tanaka, Fumiaki Yumoto, Wei Zhao, Kensuke Tono, So Iwata, James Holton, Nicholas K. Sauter, and James S. Fraser

Introduction

A major challenge for time-resolved crystallography is the development of a generalized perturbation. Currently, most experiments have focused on photoactive proteins[1–3], which can be excited by lasers emitting visible light. Given that a limited fraction of the proteome is photoactive, a general perturbation is needed. Temperature is one potential perturbation, since all proteins' conformational equilibria are affected by temperature[4,5]. In previous work[6], methods for the analysis of temperature jumps using X-ray scattering data were described. Those methods were constructed for solution scattering data, but we wished to conduct a similar experiment in the context of crystalline diffraction. Since Bragg peaks render the initial program ineffective, we needed to adapt our analysis pipeline to operate in the presence of diffraction. After adapting the method, we tested it during a time-resolved temperature-jump crystallography experiment, aiming to develop the generalized perturbation described above.

Methods

Sample preparation for serial X-ray experiments

CypA crystals were formed in batch, as described in chapter 2, at a constant stir rate of 500 RPM. Lysozyme crystals were also formed in batch. Respective batches were embedded in hydroxyethyl cellulose carrier medium. Prior to sample injection, crystal slurries were centrifuged, the supernatant was removed, and crystals were directly mixed with a 15% hydroxyethyl cellulose in a 1:9 crystal-to-cellulose ratio, as previously described[7].

Serial X-ray data collection & analysis

We collected data at SACLA[8] using a Diverse Application Platform for Hard X-ray Diffraction in SACLA (DAPHNIS)[9] at BL2[10]. For time-resolved datasets, crystals were rapidly heated by excitation of the water O-H stretch with mid-IR laser light (1443nm, 7ns pulse duration). Since protein crystals are approximately 50% solvent by volume, the IR laser effectively heated the entire volume of an irradiated crystal. At defined time delays following the IR heating pulse (20ns, 20μs, and 200μs), we probed the sample with ultrafast, high-brilliance X-ray pulses from the SACLA XFEL. Diffraction images were collected using a custom-built 4M pixel detector with multi-port CCD (mpCCD) sensors[11]. Data collection was supported by a real-time data processing pipeline[12] developed on Cheetah[13]. Images identified as hits were processed using methods from CCTBX[14]. For Bragg processing, data were indexed and integrated using *dials.stills_process*[15]. Initial indexing results were used to refine the detector model, as well as crystal models[16]. Refinement of the detector distance and panel geometry improved the agreement between measured and predicted spots. Bragg data were then merged and post-refined using *cxi.merge*[17]. Error estimates were treated according to the Ev11 method[16,18], wherein error estimates were increased using terms refined from the measurements until they could better explain the uncertainty observed in the merged reflection intensities. For diffuse processing, radial averages were calculated from each image identified as a hit using the *dxtbx.radial_average* function. The function was modified to include a sliding

one-dimensional median filter, with a window size of 15 q bins. The median filter minimized contributions of the Bragg peaks to the radial average, while preserving the diffuse scattering signal. These smoothed radial averages were dumped to a plain text file for later analysis. Images were culled from later SVD analysis based on collection anomalies (clogging of sample jet, etc.), since these contributions dominated variations in the diffuse signal, making temperature jump detection difficult.

Processing X-ray scattering curves

A series of custom scripts were written to process the radial average curves (https://github.com/fraser-lab/xray_thermometer). The flat text files dumped by CCTBX were read, and parsed into separate curves. A mask was then applied for scaling purposes, so that only data between 2.0 $Å^{-1}$ and 2.5 $Å^{-1}$ were used for scaling. All scaling of X-ray scattering curves was performed using an algebraic (least-squares) procedure, as described in Chapter 2. To determine the scale factor, A, which can be applied to a scattering curve I(q)a in order to scale it to a second scattering curve I(q)b, we used the following equation: $A = \Sigma qI(q)aI(q)b$ $\Sigma qI(q)a$.

X-ray thermometry

Following the initial data processing steps described above, we used singular value decomposition (SVD) to detect the temperature-jump introduced by the IR laser pulse[19–22]. We pooled a series of dark and light images from time-resolved datasets, scaled them to a common reference over the q=2.0-2.5 $Å^{-1}$ region, truncated the low angle signal, then performed SVD on a matrix built from these scaled curves. In this matrix, each column represents a single scattering curve, with the rows of the matrix corresponding to q-bins and the entries in the matrix consisting of azimuthally averaged scattering intensities. As described in the Results section, the SVD identified a left singular vector whose contribution to the overall scattering signal was highly temperature dependent. This was the left singular vector with the second largest singular value (U1). For each scattering curve, we calculated the average value

of $v1_n$, which is the entry in the matrix V describing the contribution of the temperature-dependent singular vector (U1) to the nth scattering curve. The values within the column in the V matrix (V1) corresponding to the temperature dependent vector (U1) allowed us to classify images as “dark” or “light”.

Results

Both CypA and Lysozyme crystals diffracted to high resolution, but only the Lysozyme data was amenable to structure solution (**Table 3.1**) for reasons previously described[23]. Briefly, variations in CypA crystal size led to overloads and streaks on the CCD detector, inhibiting proper integration of Bragg data. Nonetheless, both proteins were tested within the context of the time-resolved temperature-jump experiment described. Images were processed in real time, with temperature-jump assessment focusing upon the diffuse scattering signal within the angular region typically known as the “solvent ring” (**Figure 3.1** - top panel). By computing a radial average (**Figure 3.1** - middle panel), and using a one-dimensional median filter to remove contributions from the Bragg peaks (**Figure 3.1** - bottom panel - right side), we were able to home in on changes to the diffuse signal. Utilizing a 1D median filter was significantly faster than more traditional 2D filters (**Figure 3.1** - bottom panel - left side), which proved critical when making real-time decisions during data collection.

Data cleaning was critical for success of temperature analysis via SVD. Images were culled if an exceptional event, such as jet clogging occurred. We found that the variations in diffuse signal driven by these events could drown out any temperature dependent events. In addition to image selection, we scaled the curves based upon a narrow q range (2.0 $Å^{-1}$ to 2.5 $Å^{-1}$). This region bypassed the low angle regions that varied most greatly from changes in crystal size, laser power, and extraneous scattering mass such as fluctuations in sample jet size. The influence of Bragg peaks upon the radial average was also minimal in this angular region, mitigating any bias introduced by the median filter. Nonetheless, this angular region contained

strong diffuse scattering that is classically attributed to the carrier media, known as the “solvent ring” (**Figure 3.1**).

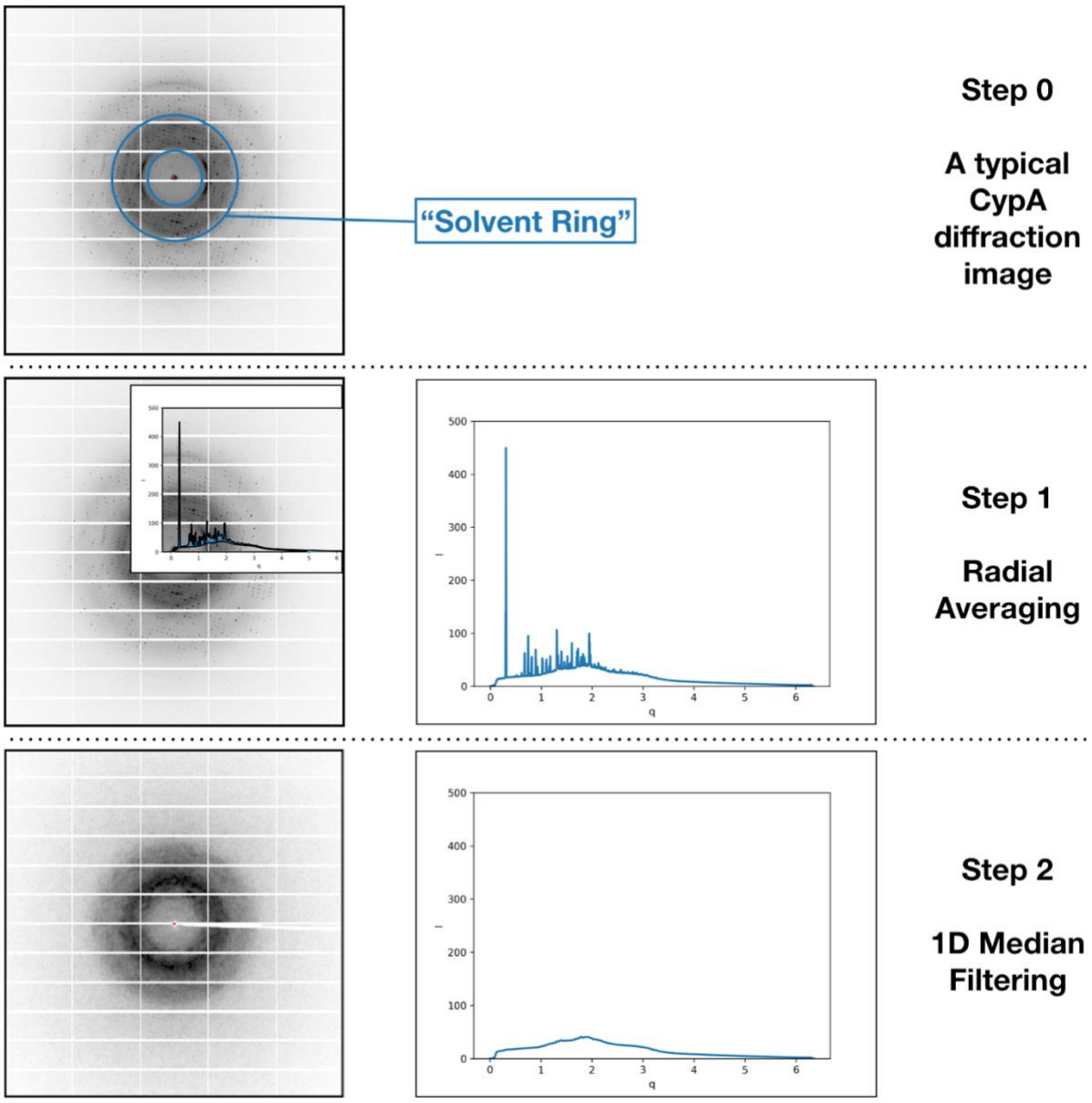

Figure 3.1 Visualization of diffraction image processing. Diffraction images contain both Bragg peaks and solvent scattering data (**step 0**). Temperature jumps primarily affect scattering within the “solvent ring,” highlighted above. To detect changes within this ring, we took radial averages (**step 1**), and then smoothed out the contributions from the Bragg peaks using a one-dimensional median filter (**step 2**). This effectively eliminates the contribution of the Bragg scattering (example of two-dimensional median filter visualized on the lower left), allowing us to focus solely on the diffuse signal contributed by the carrier media.

With a series of scaled radial averages in hand, we compiled them into a single matrix (**Figure 3.2**). Then, we decomposed this matrix using SVD, to gain insights into the main sources of variation across the images. Initial investigations focused on images within a single dataset. Data were collected in an interleaved fashion, with two “dark” images taken for every IR-illuminated “light” image. The status of the IR laser was recorded in the image headers, and data were sorted by run within Cheetah. Thus, the input to the SVD analysis was sorted first by run number, then by “dark1, dark2, light” status. This is visualized in the data structure (**Figure 3.3**), and validates the coordination of the Laser diode with the temperature dependent vector (U1).

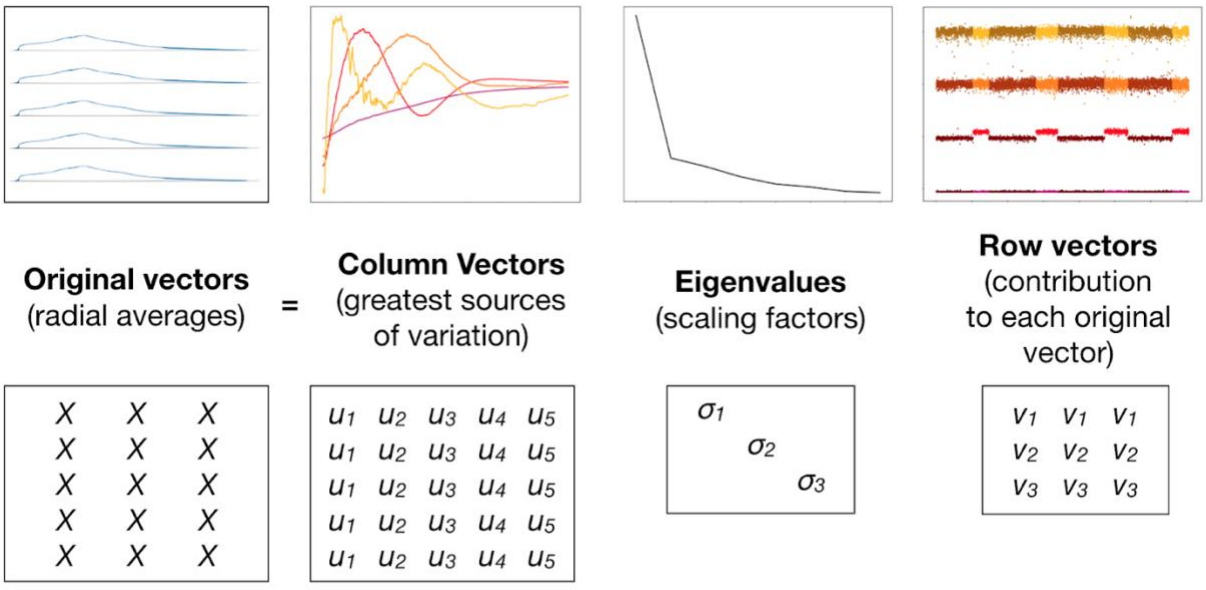

Figure 3.2 Diagram of singular value decomposition (SVD) used to detect temperature jumps. Following the preprocessing from figure 3.1, radial averages are accumulated into a matrix (**original vectors**). By decomposing this matrix using SVD, we acquire three matrices. The first (**column vectors**) reports on the vectors that maximize variation across the matrix. The intermediate matrix (**eigenvalues**) provides scaling factors for the contribution of each singular vector to the overall variation. The right singular vectors (**row vectors**)report on the per-image perturbation driven by each left singular vector (**column vectors**).

Following SVD, we examined the singular vectors associated with the largest eigenvalues (**Figure 3.3**). The strongest singular vector (U0) appears to have an exponential

decay from the origin of reciprocal space, while the next three strongest vectors (U1, U2, and U3) all appear to be approaching a decaying sinusoidal function. Smooth singular vectors are characteristic of solution scattering data, as shown in Chapter 2, so we interpreted this as a sign that the contributions of the Bragg peaks were cleanly masked by the median filter (**Figure 3.1**). Importantly, we wanted to assess the origins of this variation, so we investigated the singular values (Vn) associated with each singular vector (**Figure 3.3**). It appeared that the second strongest singular vector, U1, corresponded with the status of the IR laser.

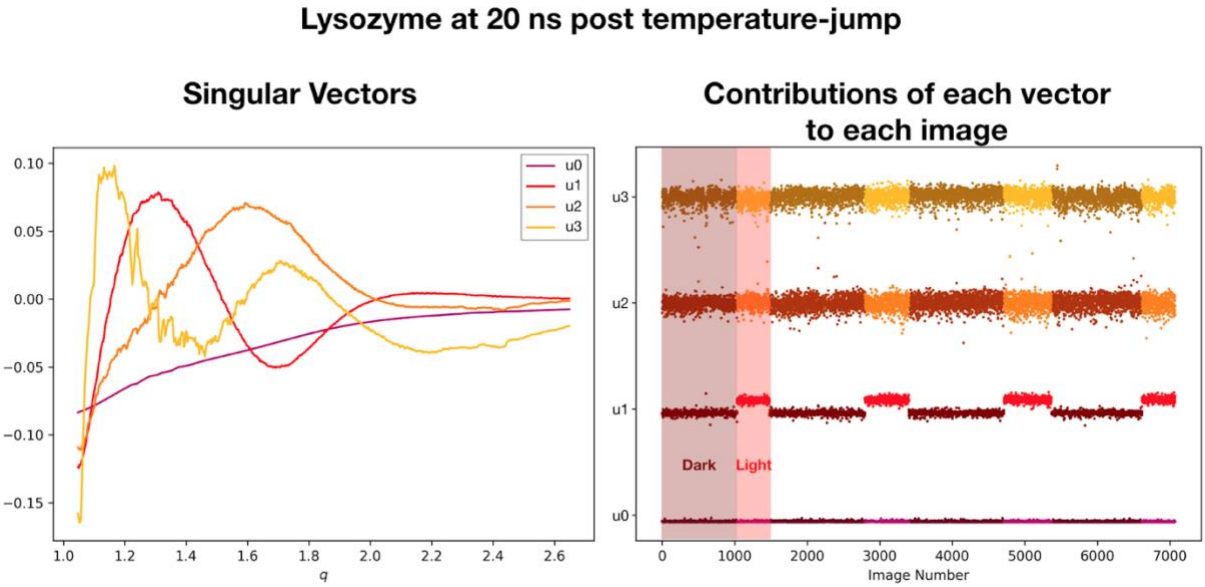

Figure 3.3 Temperature jump detection in lysozyme crystals. Following SVD analysis of over 7000 diffractions images collected during a temperature-jump experiment, the four strongest singular vectors were plotted (**left panel**). The contribution of each vector to each image is plotted (**right panel**), with “dark” and “light” images, as reported by the IR laser diode, highlighted for clarity. It appears that the second strongest vector, U1, correlates with a perturbation to the diffuse scattering that is driven by the IR laser.

To investigate this further, we plotted the singular values as histograms, capitalizing on the labels from the IR laser status to distinguish light from dark images (**Figure 3.4**). This allowed us to more carefully assess the association of any given singular vector with the IR laser status. Out of the first four singular vectors it appeared as though only U1 correlated with IR laser status, leading us to refer to U1 as the temperature-dependent vector.

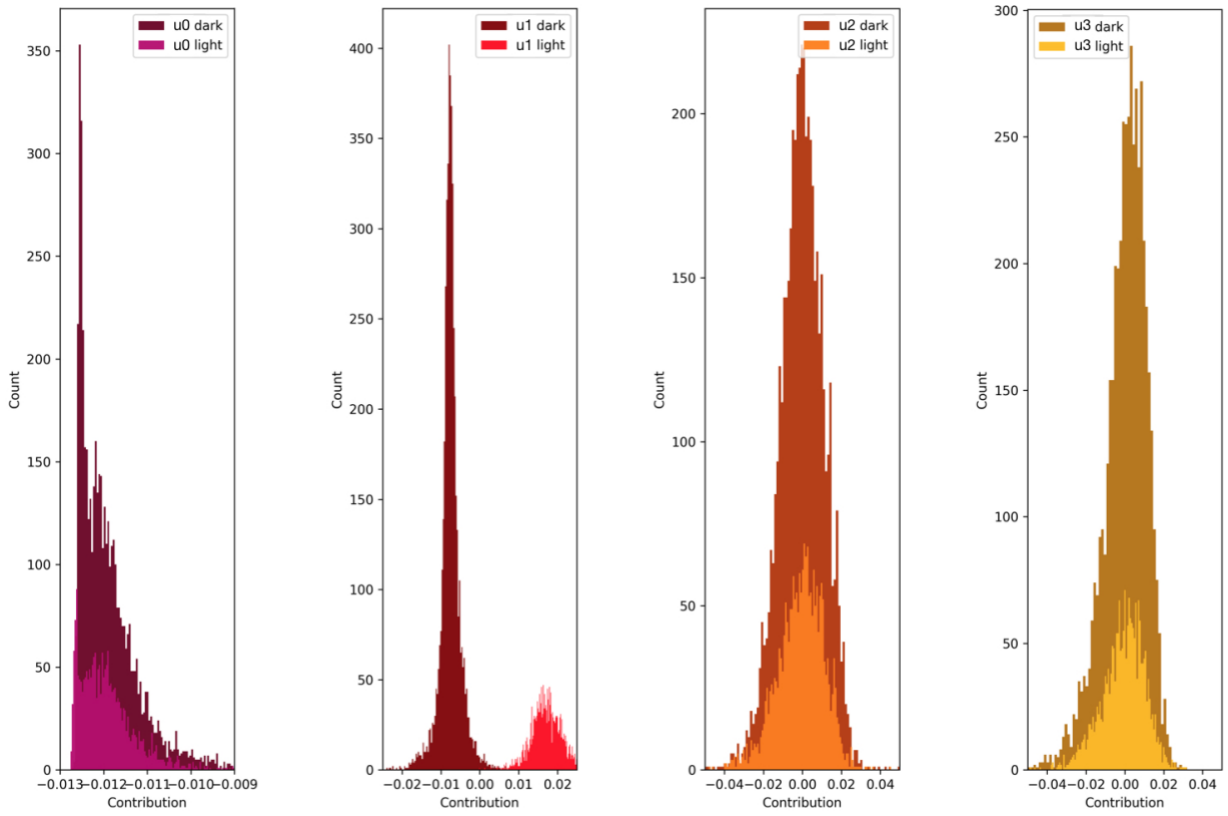

Figure 3.4 Column vector values plotted as histograms. By plotting the column vectors as histograms, clarity is gained as to whether a vector correlates with the IR laser status. This data corresponds to figure 3.3, where SVD was carried out on ~7,000 lysozyme diffraction images, ⅓ of which corresponded to a 20 ns time delay following excitation with an IR laser. In this dataset, U1 correlates with the IR laser status, allowing detection of temperature-jumps directly from diffraction data.

We initially discovered a temperature-dependent vector within one time-resolved dataset, but we wondered how this vector might vary as a function of time delay and protein characteristics. Thus, we analyzed a series of datasets, consisting of different time delays, as well as different model proteins. To our surprise, not only was there a temperature-dependent vector in each dataset, but they appeared remarkably similar (**Figure 3.5**). Since the data were scaled to a relative size prior to SVD, the amplitude of the vectors may vary on an absolute scale. What is clear is that the position of the temperature-dependent vector in angular space is consistent, as is the overall sinusoidal shape. This was consistent across time-delay and across

two different protein samples, but it did vary when an alternative carrier media was used. All datasets shown (**Figure 3.5**) were injected using cellulose carrier media, but additional datasets were solved using LCP carrier media. LCP datasets exhibited different diffuse scattering, as well as a different temperature-dependent vector. Despite this consistency, variations are still noticeable. Most prominently, the pink vector corresponding to the 20 µs Lysozyme dataset (**Figure 3.5**) displays remnants of jet clogging near 1.7 $Å^{-1}$. Large variations in diffuse scatter within a dataset that are too broad to be smoothed by the median filter can easily dominate the temperature-jump signal.

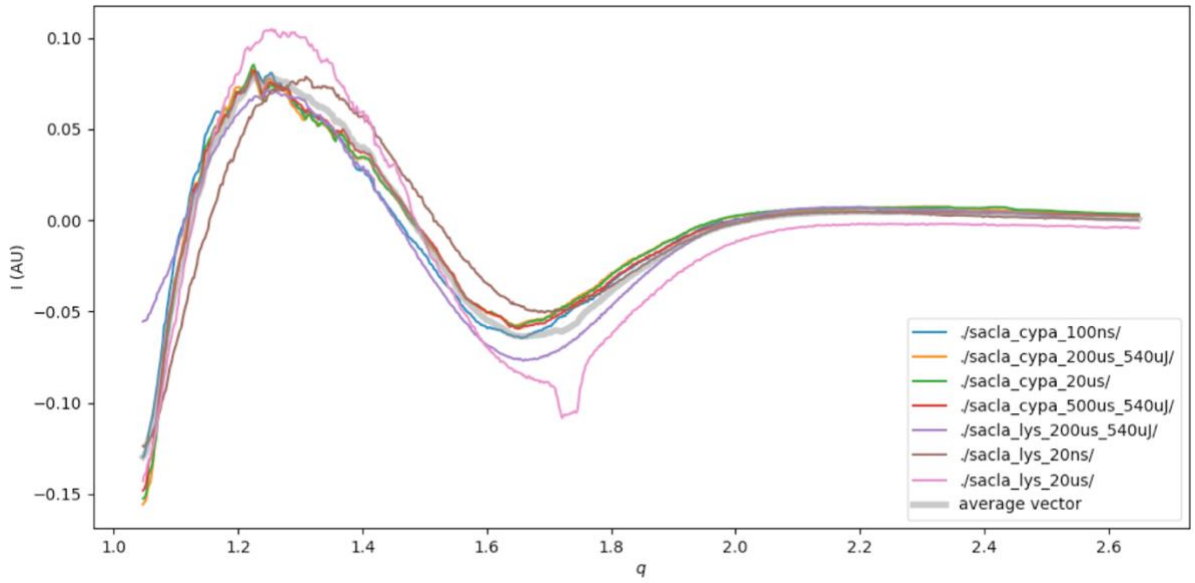

Figure 3.5 The temperature-jump signal is consistent across time delays and protein samples. The singular vector (or a linear combination of vectors) that correlated with IR laser status was plotted for seven separate datasets. Overall features and location in angular space appear to match well across the datasets, indicating that the IR-driven change in diffuse scattering is a function of the carrier media rather than of the protein crystals, or the time delay following perturbation with the IR laser.

Table 3.1. Crystallographic statistics for data collection. Statistics for the highest-resolution shell are shown in parentheses.

Lysozyme Dataset	true dark	20ns	20us	200us
X-ray Source	SACLA - BL2	SACLA - BL2	SACLA - BL2	SACLA - BL2
Photon Energy (keV)	10	10	10	10
X-ray pulse duration (fs)	~10	~10	~10	~10
Photons per pulse	~1*10 E11	~1*10 E11	~1*10 E11	~1*10 E11
Repetition Rate (Hz)	30	30	30	30
Hit rate (%)	22.4	11.6	13.1	8.9
Indexing rate (%)	84.3	88.6	87.9	82.9
No. of images	9464	15253	13681	11931
Resolution Range	28.09 - 1.57	31.62 - 1.57	28.11 - 1.57	28.11 - 1.57
Space Group	P 43 21 2	P 43 21 2	P 43 21 2	P 43 21 2
Unit-cell parameters				
a (Å)	79.45 +/- 0.10	79.52 +/- 0.13	79.52 +/- 0.10	79.51 +/- 0.11
b (Å)	79.45 +/- 0.10	79.52 +/- 0.13	79.52 +/- 0.10	79.51 +/- 0.11
c (Å)	38.16 +/- 0.07	38.23 +/- 0.08	38.21 +/- 0.05	38.19 +/- 0.05
alpha (º)	90	90	90	90
beta (º)	90	90	90	90
gamma (º)	90	90	90	90
Total Reflections	1593940 (39912)	3195992 (85262)	2208287 (56144)	1783773 (45477)
Multiplicity	90.55 (46.63)	181.56 (99.61)	125.46 (65.59)	101.34 (53.13)
Completeness (%)	99.97 (100)	99.98 (100.00)	99.97 (100.00)	99.97 (100.00)
Mean I/sigma (I)	9.463 (5.889)	12.969 (6.918)	10.844 (6.633)	9.986 (6.280)
Wilson B Factor (Å 2)	14.1	15.37	15.38	15.1
Rsplit (%)	9.4 (40.0)	5.9 (25.9)	7.5 (25.7)	8.5 (30.6)
CC int (%)	98.6 (17.0)	99.6 (34.6)	99.2 (44.5)	98.4 (37.7)
Software	cctbx + cxi.merge	cctbx + cxi.merge	cctbx + cxi.merge	cctbx + cxi.merge

With our temperature-detection method in-hand, we felt confident in proceeding with the experiment. Thus, we collected and analyzed a series of time-resolved diffraction datasets (**Table 3.1**). Each dataset was extremely high in quality, and the diffuse scattering associated with it held clear evidence of a consistent temperature-jump (**Figure 3.5**).

Discussion

By adapting methods developed for solution scattering experiments for diffraction experiments, we have enabled the first direct detection of temperature jumps from time-resolved XFEL datasets. This method allows experimentalists to verify that a temperature jump is taking place, by direct observation of the X-ray data. Furthermore, this method is compatible with online analysis, enabling near real-time feedback on the experiment as it is conducted. This is invaluable when collecting time-resolved data.

This method currently depends upon SVD to extract a temperature-dependent singular vector from the radial averages. Nonetheless, it appears that temperature-dependent vectors are consistent when carrier media and the crystalline well-solution are comparable. Thus, it may be possible to alter these methods by measuring a temperature-dependent vector prior to the time-resolved work. A temperature-dependent vector can be measured for each carrier media, and then a simple dot product of this vector with any radial average should give a value whose sign will correspond to illumination status. Alternatively, gaussian mixture models may be utilized to identify signature changes in the diffuse scattering signal. Each approach would vastly simplify the computational pipeline, increase its robustness in the presence of experimental noise, and greatly enhance the speed of analysis. Nonetheless, limitations will apply. The experimental setup must be carefully measured to ensure that the vector's position in angular space matches that of the experiment. Furthermore, a series of static measurements at defined temperatures will enable quantitative measurement of the temperature jump, as conducted in previous solution scattering experiments[6].

Finally, the most exciting application is to carry out time-resolved structural analysis, enabled by this pipeline. As shown here, we have collected a series of time-resolved datasets, and are currently investigating time-resolved structural changes. This is the ultimate goal of time-resolved structural biology, to perturb samples in a general fashion and thus to learn how each molecule populates and moves through its energy landscape.

References

1. Coquelle, N. *et al.* Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* **10**, 31–37 (2018).
2. Pande, K. *et al.* Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725–729 (2016).
3. Nogly, P. *et al.* Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* **361**, (2018).
4. Frauenfelder, H., Fenimore, P. W. & Young, R. D. Protein dynamics and function: Insights from the energy landscape and solvent slaving. *IUBMB Life* **59**, 506–512 (2007).
5. Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Parak, F. G. Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc. Natl. Acad. Sci.* **99**, 16047–16051 (2002).
6. Thompson, M. C. *et al.* Temperature-jump solution X-ray scattering reveals distinct motions in a dynamic enzyme. *Nat. Chem.* **11**, 1058–1066 (2019).
7. Sugahara, M. *et al.* Hydroxyethyl cellulose matrix applied to serial crystallography. *Sci. Rep.* **7**, 703 (2017).
8. Ishikawa, T. *et al.* A compact X-ray free-electron laser emitting in the sub-ångström region. *Nat. Photonics* **6**, 540–544 (2012).
9. Tono, K. *et al.* Diverse application platform for hard X-ray diffraction in SACLA (DAPHNIS): application to serial protein crystallography using an X-ray free-electron laser. *J. Synchrotron Radiat.* **22**, 532–537 (2015).
10. Tono, K., Hara, T., Yabashi, M. & Tanaka, H. Multiple-beamline operation of SACLA. *J. Synchrotron Radiat.* **26**, 595–602 (2019).
11. Kameshima, T. *et al.* Development of an X-ray pixel detector with multi-port charge-coupled device for X-ray free-electron laser experiments. *Rev. Sci. Instrum.* **85**, 033110 (2014).

12. Nakane, T. *et al.* Data processing pipeline for serial femtosecond crystallography at SACLA. *J. Appl. Crystallogr.* **49**, 1035–1041 (2016).
13. Barty, A. *et al.* Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Crystallogr.* **47**, 1118–1131 (2014).
14. Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W. & Adams, P. D. The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework. *J. Appl. Crystallogr.* **35**, 126–136 (2002).
15. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 85–97 (2018).
16. Brewster, A. S. *et al.* Improving signal strength in serial crystallography with DIALS geometry refinement. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 877–894 (2018).
17. Sauter, N. K. XFEL diffraction: developing processing methods to optimize data quality. *J. Synchrotron Radiat.* **22**, 239–248 (2015).
18. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292 (2011).
19. Rimmerman, D. *et al.* Direct Observation of Insulin Association Dynamics with Time-Resolved X-ray Scattering. *J. Phys. Chem. Lett.* **8**, 4413–4418 (2017).
20. Cho, H. S. *et al.* Dynamics of Quaternary Structure Transitions in R-State Carbonmonoxyhemoglobin Unveiled in Time-Resolved X-ray Scattering Patterns Following a Temperature Jump. *J. Phys. Chem. B* **122**, 11488–11496 (2018).
21. Skov Kjær, K. *et al.* Introducing a standard method for experimental determination of the solvent response in laser pump, X-ray probe time-resolved wide-angle X-ray scattering experiments on systems in solution. *Phys. Chem. Chem. Phys.* **15**, 15003–15016 (2013).
22. Cammarata, M. *et al.* Impulsive solvent heating probed by picosecond x-ray diffraction. *J. Chem. Phys.* **124**, 124504 (2006).

23. Wolff, A. M. *et al.* Comparing serial X-ray crystallography and microcrystal electron diffraction (MicroED) as methods for routine structure determination from small macromolecular crystals. *IUCrJ* **7**, 306–323 (2020).

Chapter IV

Bringing diffuse X-ray scattering into focus

Contributing Authors

Michael E Wall [1] Alexander M Wolff [2,3] James S Fraser [3]

1. Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
2. Graduate Group in Biophysics, University of California San Francisco, San Francisco, CA 94158, USA
3. Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94158, USA

Preface

The bulk of this work appears as Wall, Wolff, & Fraser, published in *Current Opinion of Structural Biology* in 2017.

Upon joining the Fraser Lab, I was initiated into the dark art of analyzing diffuse X-ray scattering. The phrase, “diffuse scattering” was rather opaque to me in the beginning. So, to shore up my understanding I was given the opportunity to collaborate with Michael Wall. Michael is a longstanding expert in the field of diffuse scattering, and he provided a great deal of mentorship during my graduate career. In the following work, we summarize the state of the art in diffuse scattering analysis as of late 2017, and provide perspective on the future growth we hope to see in the field.

Abstract

X-ray crystallography is experiencing a renaissance as a method for probing the protein conformational ensemble. The inherent limitations of Bragg analysis, however, which only reveals the mean structure, have given way to a surge in interest in diffuse scattering, which is

caused by structure variations. Diffuse scattering is present in all macromolecular crystallography experiments. Recent studies are shedding light on the origins of diffuse scattering in protein crystallography, and provide clues for leveraging diffuse scattering to model protein motions with atomic detail.

Introduction

With over 100 000 X-ray structures deposited in the wwPDB[1], improvements in data processing pipelines, and the advent of completely unattended data collection, it seems hard to imagine that there are any aspects of protein X-ray crystallography that remain to be optimized. However, only about half of the X-rays scattered by the crystalline sample are currently being analyzed — those in the Bragg peaks. The weaker, more smoothly varying features in diffraction images, known as diffuse scattering, are largely ignored by current practices. While the analysis of diffuse scattering is an established method in the fields of small molecule crystallography[2] and materials science[3], there are only very few foundational studies of diffuse scattering in macromolecular crystallography[4–18]. However, the relative scarcity of diffuse scattering studies is poised to change as activity in the field has recently increased.

A small group of researchers met in 2014 to discuss the challenges and opportunities of investigating macromolecular diffuse scattering[19]. Our attention was drawn to several key developments in the field of macromolecular crystallography that motivated and enabled assessment of the diffuse signal. First, structural models obtained using traditional methods appear to be reaching a plateau in quality, as R factors remain relatively high compared to what can be achieved in small-molecule crystallography. The origin of this ‘R-factor gap’ is likely due to the underlying inadequacies of the structural models refined against crystallographic data[20]. These inadequacies can only be overcome if we can improve the modeling, including, for example, conformational heterogeneity (especially in data collected at room temperature[21]), solvation, and lattice imperfections that break the assumptions of ‘perfect crystals’ used in data

reduction and refinement. Second, new detectors were enabling collection of data with lower noise, higher dynamic range, and highly localized signal. Third, new light sources were emerging with very bright, micro-focused beams (e.g. X-ray free-electron lasers). Collectively, these factors made us optimistic that diffuse scattering data both was needed and could be measured accurately enough to improve structural modeling. In early 2017, many of us met again to discuss the progress of the field with respect to each of these challenges identified in 2014[22]. In this review, we provide our perspective on this progress and the status of the field, informed in part by our observations at that meeting and advances covered by Meisburger et al.[23]. While there have been exciting developments in recent years, there are still major challenges ahead, include modeling atomic motions in protein crystals using diffuse scattering data with accuracy comparable to the Bragg analysis, and utilizing these models of protein motions to distinguish between competing biochemical mechanisms.

Data Collection

Extraction of diffuse scattering data from conventional protein crystallography experiments is becoming straightforward thanks to the increased accessibility of photon-counting pixel array detectors (PADs, e.g. Pilatus detectors). These detectors have greater dynamic range and do not suffer from 'blooming' overloads that obscured diffuse signals near Bragg peaks on conventional charge-coupled device (CCD) detectors. (An early CCD detector was programmed to drain excess charge away from overflowing pixels to enable measurement of diffuse scattering data[18,24]; however, this feature was not implemented in commercial detectors.) Additionally the use of PADs has led to changes in data collection strategies, such as the use of fine phi angle scans, that facilitate analysis of Bragg peaks and diffuse features from the same set of images[25]. A second major advance is the measurement of diffuse scattering using an X-ray free-electron laser (XFEL) in a serial femtosecond crystallography (SFX) experiment[26]. Using an XFEL enables collection of radiation-damage-free room

temperature data, as well the potential to examine time-resolved changes in the diffuse scattering signal.

Despite these advances in collection of diffuse scattering data, minimizing background scattering remains the most important obstacle to collecting high quality data. While it is possible to remove some background scattering during data processing, the cleanest separation requires one to remove scattering extraneous to the crystal during the experiment. Factors to consider during collection of single crystal datasets include the thickness and orientation of the loop (for relevant mounting schemes), the volume of liquid surrounding the crystal, and the amount of airspace between the crystal and the detector. Background air scatter can be also reduced by a Helium or vacuum path between sample and detector. Collection of SFX data adds additional complexity, as the injection stream and crystal size will vary. Ayyer et al.[26] addressed this challenge by selecting only the frames with the strongest diffuse scattering signal, in which the size of the crystal was expected to be comparable to the width of the jet. As the landscape of sample delivery devices for SFX and conventional crystallography continues to evolve, mounted sample delivery on materials such as graphene[27] provides a promising route for minimization of background scattering.

Data Integration

Early studies of protein diffuse scattering focused on explaining features in individual diffraction images. The introduction of methods for three-dimensional diffuse data integration enabled quantitative validation of models of correlated motions[18]. Several approaches to 3D data integration now have been implemented[24,26,28–30]. These approaches differ in several key ways: the scaling of intensities when merging the data; the handling of intensities in the neighborhood of the Bragg peak; and the strategy for sampling of reciprocal space. In the Lunus software for diffuse scattering (https://github.com/mewall/lunus) we have chosen:

(1) To use the diffuse intensity itself to scale the diffuse data (as opposed to using the Bragg peaks, as in Ref.[29]). This choice avoids artifacts due to potential differences in the way the Bragg and diffuse scattering vary with radiation damage and other confounding factors. The response of these signals to damage requires further study before a definitive scaling strategy can be chosen.

(2) To ignore or filter intensity values in regions where the variations are sharper than the 3D grid that will hold the integrated data. This can include masking halo intensities too close to a Bragg peak, and kernel-based image processing to remove Bragg peaks from diffraction images. These steps avoid the mixing of signal associated with sharp features into the signal associated with larger-scale, cloudy diffuse features. The sharply varying features (e.g. streaks) are an important component of the signal; however, to avoid artifacts in analysis, we prefer to measure them on a grid that is fine enough to resolve them[17]. If the sampling is finer than one measurement per integer Miller index, but still too coarse to resolve the halos, and if the halo intensity is nevertheless included (as in Ref.[29]) then the measurements at integer Miller indices may be segregated from the rest of the data and analyzed separately.

(3) To sample the data on a grid that includes points at Miller indices (corresponding to where the Bragg peaks are located), and, for finer sampling, points corresponding to integer subdivisions of Miller indices. Sampling strategies that are not tied to the reciprocal lattice also are valid (as used in Refs.[26,28]); however, on-lattice strategies enable leveraging of existing crystallographic analysis and modeling tools for diffuse scattering.

Efforts are now underway to decrease the burden of diffuse data integration and make diffuse data collection accessible for any protein crystallography lab. Recent algorithmic improvements have led to scalable, parallelized methods for real-time processing of single-crystal synchrotron data, decreasing the time required to extract a diffuse dataset from

diffraction images. These improvements aim to keep pace with real-time analysis of Bragg data at high frame rates, such as those expected at LCLS-II and euXFEL. Initial tests mapped staphylococcal nuclease diffuse data onto a fine-grained reciprocal lattice, using two samples per Miller index[31]. This implementation of the Lunus software is capable of processing thousands of diffraction images within a few minutes on a small computing cluster.

In addition to improving the scalability of diffuse scattering data processing, efforts are underway to create a push-button diffuse data processing pipeline. The Sematura pipeline (https://github.com/fraser-lab/diffuse_scattering) was inspired by the user-friendly environment provided by software for analyzing Bragg peaks, such as xia2[32]. To ensure portability the project was built upon the CCTBX framework[33], with future work focusing on developing Sematura as a CCTBX module for ease of access.

Building and refining models of protein motions

Liquid-like motions

After early experiments on tropomyosin[15], the liquid-like motions (LLM) model became a key tool in interpreting diffuse features in diffraction images[4,6]. In the LLM model, the crystal is treated as a soft material. All atoms are assumed to exhibit statistically identical normally distributed displacements about their mean position. The correlation between atom displacements is a decreasing function of the distance between the atoms, usually an exponential decay. A LLM was used to interpret early 3D diffuse data sets, refined using a correlation coefficient as a target function[18]. Successful refinement of a LLM model was used to demonstrate that interpretable diffuse datasets can be extracted from Bragg diffraction experiments, when data collection is not specifically targeted at measuring the diffuse signal[25]. Peck et al.[29] recently found the ability of the LLM to include correlations across unit cell boundaries was essential for modeling the diffuse signal in several 3D datasets. This result is intuitive, as many atoms in a typical protein crystal are within 5–10 Å of symmetry related

molecules. Overall the LLM model has proven to be a simple means of explaining the general features of the data with a straightforward interpretation, and therefore remains an important first approach to analysis of protein diffuse X-ray scattering data.

Normal mode analysis and elastic network models

Beyond the LLM model, normal mode analysis (NMA) of elastic network models (ENMs) can provide insights into the soft modes of protein dynamics in more detail, helping to reveal mechanisms that bridge protein structure and function[34]. In an ENM, the atoms of the crystal structure are connected by springs, and the resulting network is coupled to a thermal bath. NMA then yields the covariance matrix of atom displacements. The diagonal elements of the covariance matrix correspond to the crystallographic B factors. ENMs are often used to predict B factors, which come from the Bragg analysis through the crystal structure model. However, Riccardi et al.[35] showed how to renormalize the entire covariance matrix using the crystallographic B factors. Importantly, this renormalization enables any ENM to be entirely consistent with the Bragg data, while preserving differences in diffuse scattering. Diffuse scattering could help differentiate between these ENMs because the off-diagonal elements directly influence the diffuse signal. Thus, there is an opportunity for carefully measured diffuse data to be used in refinement of ENM models, and subsequent refinement of models of protein structure and dynamics.

Indeed, many key elements needed for refinement of normal modes models using diffuse scattering already have been demonstrated. Cloudy diffuse features in X-ray diffraction from lysozyme crystals resemble the diffuse scattering predicted from simulations of normal modes models[9,13]. Similarly, sharper diffuse features in the neighborhood of Bragg peaks in ribonuclease crystals can be captured by lattice normal modes[36]. Different varieties of ENMs for staphylococcal nuclease give rise to distinct diffuse scattering patterns, even when renormalized using the crystallographic B factors[35].

Three-dimensional diffuse scattering data from trypsin and proline isomerase (CypA) recently were modeled using ENMs[25]. The agreement was substantial, considering that the models were not refined. On the other hand, Peck et al.[29] found a low agreement between ENM models and diffuse data. How much can refinement improve the agreement of an ENM model? Here we provide an example. In our example, the asymmetric unit of PDB ID 4WOR was expanded to the P1 unit cell, and an ENM was constructed as in Ref.[25]. The spring force constants between C-alpha atoms were computed as $e^{-r_{ij}/\lambda}$, where r_{ij} is the closest distance between atoms i and j, either in the same unit cell or in neighboring unit cells of the crystal structure. All atoms on the same residue as the C-alpha were assumed to move rigidly as a unit. The initial value $\lambda = 10.5$ Å yielded a linear correlation of 0.07 with the anisotropic component of the diffuse data, as computed in Ref.[25]. Powell minimization using the scipy.optimize.minimize method was used to refine the value of λ, using the negative correlation as a target. The final correlation was 0.54 for a value $\lambda = 0.157$ Å — a substantial improvement, but one that indicates that the direct interactions are essentially limited to nearest atomic neighbors. Simulated diffuse intensity in diffraction images calculated using the model vs. the data show similarities in cloudy diffuse features (**Figure 4.2**). Key strategies for improving the model are: extending from a C-alpha network to an all-atom network; using crystalline normal modes that extend beyond a single unit cell (prior studies used the Born von Karman method to compute these modes[35,36], but did not fully include the resulting modes in the thermal diffuse scattering calculation[37]); and allowing spring constants to deviate locally from the exponential behavior. Optimizing this type of model has applications beyond diffuse scattering validation and model refinement, as structures derived from normal modes analysis of network models have been useful for providing alternative starting points for molecular replacement[38] and have recently been used in an exciting local refinement procedure in cryo electron microscopy[39].

Ensemble refinement

A great promise of diffuse scattering is the potential to validate ensemble or multiconformer models of protein structures (**Figure 4.1**). As for as with Translation-Libration-Screw (TLS) refinement[40] and ENM models, diffuse signal might differ for ensembles that result in the same average structure. Even if information about atomistic conformations remains out of reach, the signal could potentially be leveraged to improve ensemble models derived from time-averaged refinement using the scheme by Gros and colleagues[41]. Currently, this procedure operates on the rationale that large scale deviations can be modeled using a TLS model, and the residual local deviations are then sampled by a molecular dynamics simulation with a time-averaged difference electron density term. Our work has revealed that diffuse scattering calculated from TLS models of disorder do not match the measured diffuse signal, however, indicating that TLS is a poor descriptor of the disorder within the protein crystals we considered[25]. Given the improvements seen when including neighboring unit cells in LLM models[29], the disorder of the crystal environment might be better accounted for by a coarse-grained model of intramolecular motion using a NMA model refined against the diffuse scattering signal. In addition, due to the limited number of copies in ensemble models, they can exhibit artificially long length scales compared to molecular dynamics simulations, which contain orders of magnitude more finely time-sliced 'snapshots' of the system[42]. Ensemble models of diffuse scattering data will therefore need to include the effect of decoherence corresponding to the finer scale motions that are filtered out in conformational selection. Once large-scale disorder is accounted for by NMA, local anharmonic deviations from the modes can be explored using MD simulations restrained by the X-ray data. As diffuse analysis becomes more sensitive, the selection of the final representative ensemble also might be optimized against the diffuse data. This selection step could supplement the current practice of selecting an ensemble that matches the Bragg data.

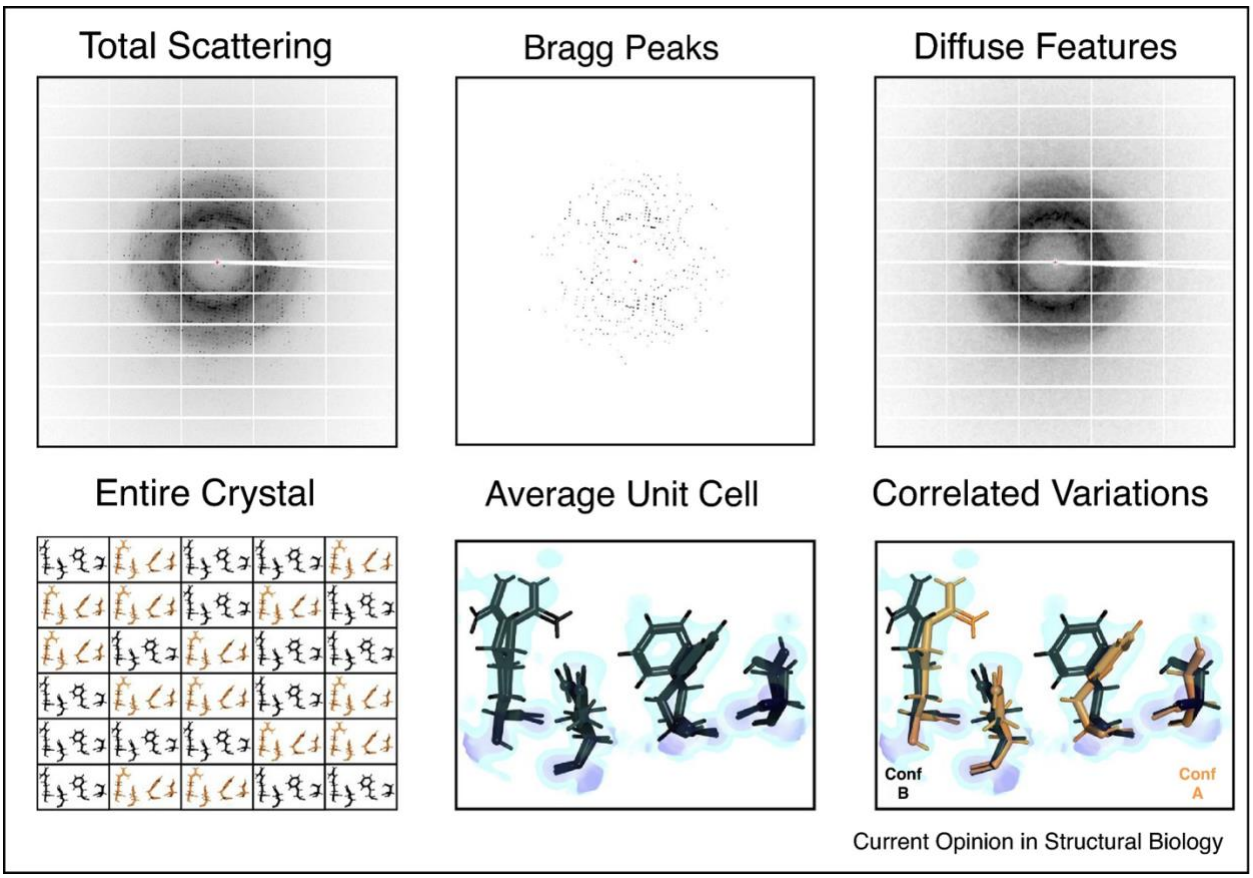

Figure 4.1. **Breakdown of Bragg and diffuse scattering**. A typical detector image in X-ray crystallography (from[25]) (upper, left) records all of the X-rays scattered by a protein crystal during a single exposure. Dark pixels correspond to high X-ray intensities. A cartoon crystal is depicted (lower, left) that contains a series of unit cells, with the contents of any given unit cell adopting one of two conformations (the conformations are expected to be more varied in a real protein crystal). Conformation A is shown in orange, while conformation B is shown in black (lower panel). During analysis, data are reduced by examining only the Bragg peaks (upper, middle), which report on the average charge density within a unit cell (lower, middle). The electron density is shown in blue, with areas of especially strong charge highlighted in purple. While multiple conformations may be modeled into the average density, assigning which conformations occur together across residues requires additional information. Current modeling practices use geometric constraints to help classify different alternative conformation groups. The diffuse scattering left behind during data reduction (upper, right) is an additional potential source of such information. Diffuse scattering includes an isotropic component that is determined both by protein and solvent scattering[43,44], and an anisotropic component that is dominated by correlated protein motions within the crystal[44]. Analyzing this anisotropic signal might help to distinguish networks of residues that move together (lower, right).

Molecular dynamics simulations

In addition to refining models of protein motions, diffuse scattering can be used to validate MD simulations[7,9,31,43–47]. Early efforts were hindered by the use of 10 ns or shorter

simulation durations[7,9,43,45], which lacked sufficient sampling for the calculations. Microsecond duration simulations of protein crystals are now becoming routine[31,44,48,49]. For staphylococcal nuclease, microsecond simulations overcome the sampling limitations for diffuse scattering calculations, while providing insight into ligand binding and catalysis[44].

The agreement of the total diffuse intensity with MD simulations is high for staphylococcal nuclease[44,46], yielding a linear correlation of 0.94 for a microsecond simulation[44]. Agreement with the 10-fold weaker anisotropic component is lower[31,44], but is more sensitive to the details of the simulation, creating opportunities for increasing the accuracy of MD models. Expanding the staphylococcal nuclease model from a single periodic unit cell to a 2 × 2 × 2 supercell increased the correlation with the anisotropic component to 1.6 Å resolution from 0.42 to 0.68 for a microsecond simulation[31].

Even though MD simulations provide a picture of crystalline dynamics at atomic detail, the accuracy of the published MD models is not yet high enough to validate the atomic details using diffuse scattering. Although we do not know what level of accuracy will be required for diffuse scattering to reveal the atomic details of protein motions, the possibility of validating the atomic details of protein motions strongly motivates improving the MD models. Ideas for improving the MD model include: increasing the size of the supercell even further, to 3x3x3 or beyond; improving force fields; increasing the simulation duration; and introducing crystal imperfections such as vacancies (missing copies of the protein) or dislocations. It is also possible that higher quality experimental data would be required to improve the model. Additional insights for increasing model accuracy might come from solid state NMR (ssNMR) experiments combined with crystalline protein simulations[50–52], which create opportunities for joint validation of MD simulations using crystallography and NMR.

Phasing and resolution extension

In a high-profile publication, the Chapman and Fromme groups integrated the first three-dimensional diffuse scattering dataset from a serial femtosecond protein crystallography experiment at an X-ray free electron laser[26]. Their analysis focused on the potential for phasing and resolution extension of a charge density map of photosystem II (PSII). In this study, the method, based on the difference-map algorithm[53], depends critically on the assumption that the diffuse signal is proportional to the molecular transform of the PSII dimer. In this respect, the work is closely related to that of Stroud and Agard[54] and Makowski [55] on phasing using continuous diffraction data. In Ref.[26], and in a follow-on study using improved data integration[56], the validity of the method was argued by assuming independent, rigid body motions of the dimer.

Using diffuse scattering for phasing and resolution extension might prove to be useful in rare cases when diffuse scattering extends to higher resolution than the Bragg diffraction; however, many technical questions remain both about the origin of diffuse scattering in PSII and the role of diffuse scattering in yielding the resulting charge density. What effect does the presence of Bragg peaks have on phasing and resolution extension in the 4.5–3.5 Å range, which is where the diffuse intensity was measured? How does the improvement in the PSII map compare to what would be obtained by using randomized intensities, due to the free lunch effect[57] and solvent flattening[58]? How would the R-factors in the extended resolution range reported in PDB 5E79 compare to pseudo-crystallographic refinement[59] of using either random intensities or the uniform average intensity in these bins? How robust are the improved features of the charge density in Ref.[26] to omit map analysis[60], especially at the solvent/protein interface? Might a LLM model (or a ENM or MD model) more accurately describe the diffuse scattering than rigid-body translations of PSII dimers (in Ref.[56], the agreement was improved when the intensities were convoluted with a 4 × 4 × 4 voxel kernel)? Can the model be improved by assuming the rigid units are coupled instead of independent[8], or if the model included rotations

as well as translations[14] (in Ref.[56], the intensities were rotationally blurred, but this does not correspond to rigid-body rotations[61])? What is the role of substitution disorder[62] (e.g. unit cells in which one or more copies of the PSII dimer are missing) in determining the diffuse signal? Understanding the implications of this method for protein crystallography will rely on answering these questions and determining whether signals at higher angle than Bragg are commonly observed.

Future perspective

The massive investment in structural genomics in the 2000s dramatically increased the robustness of X-ray crystallography data collection, processing, and refinement. The resulting technological improvements and standardizations have led to more robust methods and instrumentation for data collection that are well-aligned with the requirements for diffuse scattering experiments, enabling measurement of diffuse scattering data from traditional crystallography experiments[25,29]. These advances, along with software that makes the data processing and analysis more accessible, will enable diffuse scattering studies at any modern beamline, by any crystallography lab. Why should crystallographers take advantage of this offering? Diffuse features in protein crystallography can be myriad and complex: a mixture streaks, satellite reflections, isotropic scattering, cloudy patterns, circles, and others. Sometimes the diffuse signal appears to reflect imperfections in crystal packing, and offers a possible explanation for why a structure cannot be solved. In other cases, the diffuse signal might be so complicated as to be uninterpretable, or might be so weak that it is difficult to learn anything interesting.

Integrating diffuse scattering with Bragg diffraction to improve crystallographic models could become a major application[17,30,63]. Although assuming proteins are rigid provides the greatest potential for phasing using diffuse scattering data[26,56], multiple studies of both Bragg and diffuse scattering point to a more dynamic picture of crystalline proteins. A model with

internal motions such as the LLM tends to obscure the molecular transform signal and to limit the information to what is available from the crystal transform, at Miller indices[29]. Nevertheless, because the diffuse signal can extend well beyond the resolution limit of the Bragg peaks in rare cases, it still allows for the possibility of resolution extension. The blurring of the signal implied by the LLM means there is a loss of information in the diffuse Patterson function at long distances, so the path to resolution extension might require model refinement in addition to, or instead of, direct methods. In addition, the apparent success of the LLM[4,6,17,18,25,29] and MD simulations[31,43,44,46,47] in obtaining insights into diffuse scattering data points to a picture in which internal motions are important. This opens up the possibility that diffuse scattering can be used to reveal atomic models of protein motions, a possibility that is eliminated when proteins are treated as rigid units.

Ironically, the strongly diffracting model systems that have enabled experimental measurements of diffuse scattering may contain less informative signals than more poorly ordered crystals. In the case of poorly ordered systems, multiple significant conformational minima may co-exist in the crystal. This disorder would limit the power of Bragg diffraction, while the correlations present in the ensemble or the spread of conformations between extremes[64] would lead to diffuse signal. However, increased disorder, which may or may not be biochemically meaningful, may also present additional challenges in processing the diffuse scattering data. Nevertheless, a small but growing number of systems have shown simpler patterns of strong diffuse features that appear to be connected to protein motions. Modeling of diffuse scattering for such systems has improved substantially in recent years. For some models, like normal modes, the agreement with the diffuse signal is still relatively weak (**Figure 4.2**); for others, such as MD, the agreement is stronger[31], but still lower than what is typical in Bragg analysis.

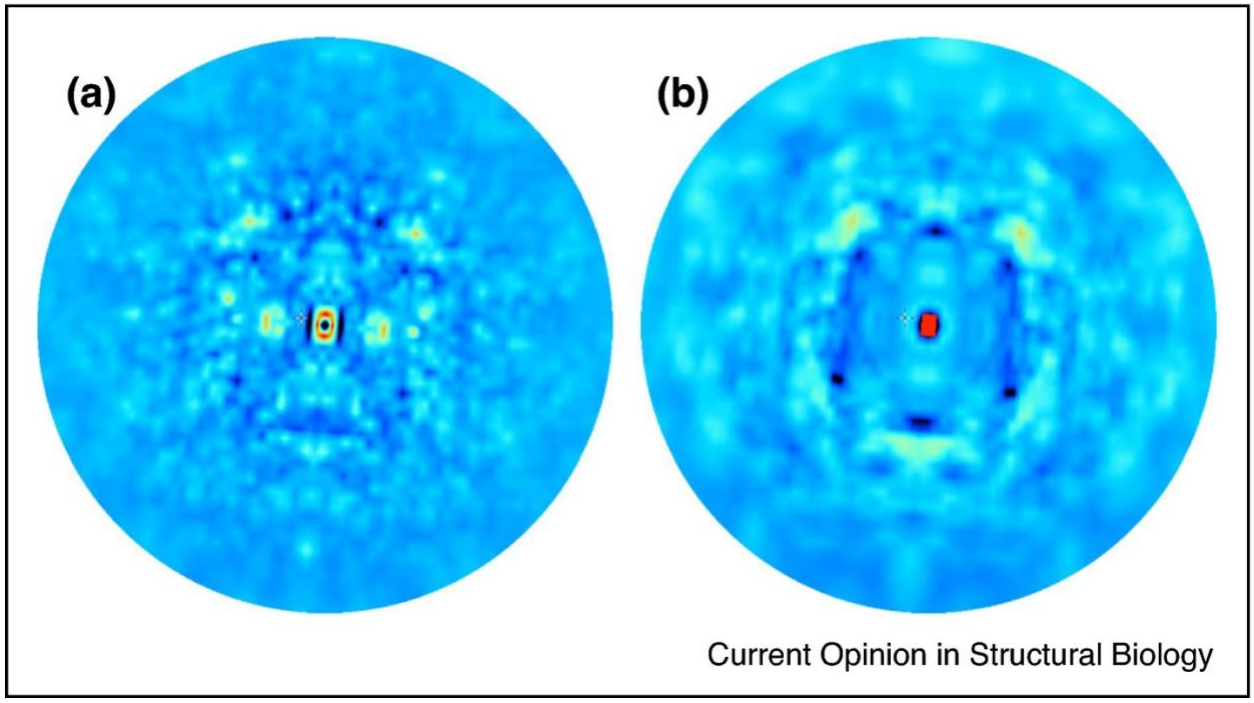

Figure 4.2. **Comparison of simulated vs experimental diffuse data.** (a) a refined ENM of staphylococcal nuclease and (b) experimental data from Ref.[18].

What will be required to take diffuse scattering to the next level and make it an equal partner to Bragg analysis? We can look to traditional crystallography for clues. In Bragg analysis, using model phases, the details of the atomic structure only are revealed when the model of the whole system is sufficiently accurate. As the accuracy is increased, the effect of refining the model of small numbers of atoms can be seen in the analysis. When we have a sufficiently accurate model for all the diffracted X-rays, diffuse scattering might become sensitive to atomic details of correlated variations, with the caveat that the diffuse signal might contain less information than the Bragg diffraction, as measurements nearby in reciprocal space can be correlated. If atomic details of correlated motions can be revealed for the systems that have the clearest diffuse signal, diffuse scattering would then begin to provide very interesting information for these systems. As capabilities evolve, more complex patterns might be tackled, and the analysis methods applied more generally. Meanwhile, it will be important to scour the diffraction image repositories such as SBGrid Databank[65] or CXIdb[66] and archive raw data from

new experiments, identifying cases where the diffuse signal is strong and amenable to analysis[67].

Despite being present in all macromolecular diffraction patterns, the origins of diffuse scattering in protein crystallography are in many cases still mysterious. Our current short-term outlook is that, for a small number of cases, whether it is due to long-range[26] or short-range disorder[25,29,31], diffuse scattering will provide valuable information for structural modeling. The types of conformational heterogeneity that can be validated and, potentially, refined against diffuse scattering data can guide us to define better models of protein structure and dynamics. As the structural biology toolkit expands, X-ray scattering, including diffuse scattering, still provides unique capabilities to probe conformational ensembles over many length scales, as captured in the recent review by Meisburger et al.[23]. Ultimately, the better models of concerted motions will have far ranging impact beyond the average structure that is accessible using conventional X-ray crystallography and cryo-electron microscopy data, yielding a deeper understanding of biochemical mechanism[68].

Conflict of interest

The authors declare there is no conflict of interest.

Acknowledgements

We thank TJ Lane, R Stroud, H Chapman, and K Ayyer, GN Phillips, Jr, and anonymous reviewers for helpful comments on the preprint. JSF is supported by a Pew Scholar Award from the Pew Charitable Trusts, a Packard Fellowship from the David and Lucile Packard Foundation, NSF STC-1231306, NIH GM123159, and NIH GM124149. JSF and MEW are supported by UC Office of the President Laboratory Fees Research Program LFR-17-476732. MEW is supported by US Department of Energy, via the Exascale Computing Project.

References

1. Burley, S. K. *et al.* Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. in *Protein Crystallography: Methods and Protocols* (eds. Wlodawer, A., Dauter, Z. & Jaskolski, M.) 627–641 (Springer, 2017). doi:10.1007/978-1-4939-7000-1_26.
2. Welberry, T. R. *Diffuse X-Ray Scattering and Models of Disorder*. (OUP Oxford, 2004).
3. Keen, D. A. & Goodwin, A. L. The crystallography of correlated disorder. *Nature* **521**, 303–309 (2015).
4. Caspar, D. L. D., Clarage, J., Salunke, D. M. & Clarage, M. Liquid-like movements in crystalline insulin. *Nature* **332**, 659–662 (1988).
5. Chacko, S. & Phillips, G. N. Diffuse x-ray scattering from tropomyosin crystals. *Biophys. J.* **61**, 1256–1266 (1992).
6. Clarage, J. B., Clarage, M. S., Phillips, W. C., Sweet, R. M. & Caspar, D. L. D. Correlations of atomic movements in lysozyme crystals. *Proteins Struct. Funct. Bioinforma.* **12**, 145–157 (1992).
7. Clarage, J. B., Romo, T., Andrews, B. K., Pettitt, B. M. & Phillips, G. N. A sampling problem in molecular dynamics simulations of macromolecules. *Proc. Natl. Acad. Sci.* **92**, 3288–3292 (1995).
8. Doucet, J. & Benoit, J. P. Molecular dynamics studied by analysis of the X-ray diffuse scattering from lysozyme crystals. *Nature* **325**, 643–646 (1987).
9. Faure, P. *et al.* Correlated intramolecular motions and diffuse x–ray scattering in lysozyme. *Nat. Struct. Mol. Biol.* **1**, 124–128 (1994).
10. Glover, I. D., Harris, G. W., Helliwell, J. R. & Moss, D. S. The variety of X-ray diffuse scattering from macromolecular crystals and its respective components. *Acta Crystallogr. B* **47**, 960–968 (1991).
11. Helliwell, J. R., Glover, I. D., Jones, A., Pantos, E. & Moss, D. S. Protein dynamics: use of computer graphics and protein crystal diffuse scattering recorded with synchrotron X-

radiation. *Biochem. Soc. Trans.* **14**, 653–655 (1986).

12. Kolatkar, A. R., Clarage, J. B. & Phillips, G. N. Analysis of diffuse scattering from yeast initiator tRNA crystals. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 210–218 (1994).
13. Mizuguchi, K., Kidera, A. & Gō, N. Collective motions in proteins investigated by X-ray diffuse scattering. *Proteins Struct. Funct. Bioinforma.* **18**, 34–48 (1994).
14. Pérez, J., Faure, P. & Benoit, J.-P. Molecular Rigid-Body Displacements in a Tetragonal Lysozyme Crystal Confirmed by X-ray Diffuse Scattering. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 722–729 (1996).
15. Phillips, G. N., Fillers, J. P. & Cohen, C. Motions of tropomyosin. Crystal as metaphor. *Biophys. J.* **32**, 485–502 (1980).
16. Polikanov, Y. S. & Moore, P. B. Acoustic vibrations contribute to the diffuse scatter produced by ribosome crystals. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 2021–2031 (2015).
17. Wall, M. E., Clarage, J. B. & Phillips, G. N. Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. *Structure* **5**, 1599–1612 (1997).
18. Wall, M. E., Ealick, S. E. & Gruner, S. M. Three-dimensional diffuse x-ray scattering from crystals of Staphylococcal nuclease. *Proc. Natl. Acad. Sci.* **94**, 6180–6184 (1997).
19. Wall, M. E., Adams, P. D., Fraser, J. S. & Sauter, N. K. Diffuse X-Ray Scattering to Model Protein Motions. *Structure* **22**, 182–184 (2014).
20. Holton, J. M., Classen, S., Frankel, K. A. & Tainer, J. A. The R-factor gap in macromolecular crystallography: an untapped potential for insights on accurate structures. *FEBS J.* **281**, 4046–4060 (2014).
21. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci.* **108**, 16247–16252 (2011).
22. Wall, M. E. (ORCID:000000031000688X). *Measurement and Interpretation of Diffuse Scattering in X-Ray Diffraction for Macromolecular Crystallography*. https://www.osti.gov/biblio/1400134 (2017) doi:10.2172/1400134.

23. Meisburger, S. P., Thomas, W. C., Watkins, M. B. & Ando, N. X-ray Scattering Studies of Protein Structural Dynamics. *Chem. Rev.* **117**, 7615–7672 (2017).
24. Wall, M. E. Diffuse Features in X-ray Diffraction from Protein Crystals. 181.
25. Benschoten, A. H. V. *et al.* Measuring and modeling diffuse scattering in protein X-ray crystallography. *Proc. Natl. Acad. Sci.* **113**, 4069–4074 (2016).
26. Ayyer, K. *et al.* Macromolecular diffractive imaging using imperfect crystals. *Nature* **530**, 202–206 (2016).
27. Sui, S. *et al.* Graphene-based microfluidics for serial crystallography. *Lab. Chip* **16**, 3082–3096 (2016).
28. Estermann, M. A. & Steurer, W. Diffuse scattering data acquisition techniques. *Phase Transit.* **67**, 165–195 (1998).
29. Peck, A., Poitevin, F. & Lane, T. J. Intermolecular correlations are necessary to explain diffuse scattering from protein crystals. *IUCrJ* **5**, 211–222 (2018).
30. Wall, M. E. Methods and Software for Diffuse X-Ray Scattering from Protein Crystals. in *Micro and Nano Technologies in Bioanalysis: Methods and Protocols* (eds. Foote, R. S. & Lee, J. W.) 269–279 (Humana Press, 2009). doi:10.1007/978-1-59745-483-4_17.
31. Wall, M. E. Internal protein motions in molecular-dynamics simulations of Bragg and diffuse X-ray scattering. *IUCrJ* **5**, 172–181 (2018).
32. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).
33. Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W. & Adams, P. D. The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework. *J. Appl. Crystallogr.* **35**, 126–136 (2002).
34. Haliloglu, T. & Bahar, I. Adaptability of protein structures to enable functional interactions and evolutionary implications. *Curr. Opin. Struct. Biol.* **35**, 17–23 (2015).
35. Riccardi, D., Cui, Q. & Phillips, G. N. Evaluating Elastic Network Models of Crystalline

Biological Molecules with Temperature Factors, Correlated Motions, and Diffuse X-Ray Scattering. *Biophys. J.* **99**, 2616–2625 (2010).

36. Meinhold, L., Merzel, F. & Smith, J. C. Lattice Dynamics of a Protein Crystal. *Phys. Rev. Lett.* **99**, 138101 (2007).
37. James, R. W. *The optical principles of the diffraction of X-rays*. (G.Bell and Sons, 1954).
38. Suhre, K. & Sanejouand, Y.-H. ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **32**, W610–W614 (2004).
39. Schilbach, S. *et al.* Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204–209 (2017).
40. Van Benschoten, A. H. *et al.* Predicting X-ray diffuse scattering from translation–libration–screw structural ensembles. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 1657–1667 (2015).
41. Burnley, B. T., Afonine, P. V., Adams, P. D. & Gros, P. Modelling dynamics in protein crystal structures by ensemble refinement. *eLife* **1**, e00311 (2012).
42. Levin, E. J., Kondrashov, D. A., Wesenberg, G. E. & Phillips, G. N. Ensemble Refinement of Protein Crystal Structures: Validation and Application. *Structure* **15**, 1040–1052 (2007).
43. Meinhold, L. & Smith, J. C. Correlated Dynamics Determining X-Ray Diffuse Scattering from a Crystalline Protein Revealed by Molecular Dynamics Simulation. *Phys. Rev. Lett.* **95**, 218103 (2005).
44. Wall, M. E. *et al.* Conformational dynamics of a crystalline protein from microsecond-scale molecular dynamics simulations and diffuse X-ray scattering. *Proc. Natl. Acad. Sci.* **111**, 17887–17892 (2014).
45. Héry, S., Genest, D. & Smith, J. C. X-ray diffuse scattering and rigid-body motion in crystalline lysozyme probed by molecular dynamics simulation11Edited by R. Huber. *J. Mol. Biol.* **279**, 303–319 (1998).
46. Meinhold, L. & Smith, J. C. Fluctuations and Correlations in Crystalline Protein Dynamics: A

Simulation Analysis of Staphylococcal Nuclease. *Biophys. J.* **88**, 2554–2563 (2005).

47. Meinhold, L. & Smith, J. C. Protein dynamics from X-ray crystallography: Anisotropic, global motion in diffuse scattering patterns. *Proteins Struct. Funct. Bioinforma.* **66**, 941–953 (2007).
48. Janowski, P. A., Cerutti, D. S., Holton, J. & Case, D. A. Peptide Crystal Simulations Reveal Hidden Dynamics. *J. Am. Chem. Soc.* **135**, 7938–7948 (2013).
49. Janowski, P. A., Liu, C., Deckman, J. & Case, D. A. Molecular dynamics simulation of triclinic lysozyme in a crystal lattice. *Protein Sci.* **25**, 87–102 (2016).
50. Kurauskas, V. *et al.* Slow conformational exchange and overall rocking motion in ubiquitin protein crystals. *Nat. Commun.* **8**, 1–12 (2017).
51. Ma, P. *et al.* Observing the overall rocking motion of a protein in a crystal. *Nat. Commun.* **6**, 1–10 (2015).
52. Mollica, L. *et al.* Atomic-Resolution Structural Dynamics in Crystalline Proteins from NMR and Molecular Simulation. *J. Phys. Chem. Lett.* **3**, 3657–3662 (2012).
53. Elser, V. & Millane, R. P. Reconstruction of an object from its symmetry-averaged diffraction pattern. *Acta Crystallogr. A* **64**, 273–279 (2008).
54. Stroud, R. M. & Agard, D. A. Structure determination of asymmetric membrane profiles using an iterative Fourier method. *Biophys. J.* **25**, 495–512 (1979).
55. Makowski, L. The use of continuous diffraction data as a phase constraint. I. One-dimensional theory. *J. Appl. Crystallogr.* **14**, 160–168 (1981).
56. Chapman, H. N. *et al.* Continuous diffraction of molecules and disordered molecular crystals. *J. Appl. Crystallogr.* **50**, 1084–1103 (2017).
57. Caliandro, R. *et al.* Phasing at resolution higher than the experimental resolution. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 556–565 (2005).
58. Wang, B.-C. Resolution of phase ambiguity in macromolecular crystallography. in *Methods in Enzymology* vol. 115 90–112 (Academic Press, 1985).

59. Fischer, N. *et al.* Structure of the E. coli ribosome–EF-Tu complex at <3 Å resolution by C s -corrected cryo-EM. *Nature* **520**, 567–570 (2015).
60. Bhat, T. N. & Cohen, G. H. OMITMAP: An electron density map suitable for the examination of errors in a macromolecular model. *J. Appl. Crystallogr.* **17**, 244–248 (1984).
61. Moore, P. B. On the Relationship between Diffraction Patterns and Motions in Macromolecular Crystals. *Structure* **17**, 1307–1315 (2009).
62. Guinier, A. *X-ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies*. (Courier Corporation, 1994).
63. Clarage, J. B. & Phillips, G. N. Analysis of diffuse scattering and relation to molecular motion. in *Methods in Enzymology* vol. 277 407–432 (Academic Press, 1997).
64. Chaudhry, C., Horwich, A. L., Brunger, A. T. & Adams, P. D. Exploring the Structural Dynamics of the E.coli Chaperonin GroEL Using Translation-libration-screw Crystallographic Refinement of Intermediate States. *J. Mol. Biol.* **342**, 229–245 (2004).
65. Meyer, P. A. *et al.* Data publication with the structural biology data grid supports live analysis. *Nat. Commun.* **7**, 1–12 (2016).
66. Maia, F. R. N. C. The Coherent X-ray Imaging Data Bank. *Nat. Methods* **9**, 854–855 (2012).
67. Kroon-Batenburg, L. M. J., Helliwell, J. R., McMahon, B. & Terwilliger, T. C. Raw diffraction data preservation and reuse: overview, update on practicalities and metadata requirements. *IUCrJ* **4**, 87–99 (2017).
68. Shoemaker, S. C. & Ando, N. X-rays in the Cryo-EM Era: Structural Biology's Dynamic Future. *Biochemistry* **57**, 277–285 (2018).

Chapter V

Optimizing diffuse scattering data collection

Contributing Authors

Alexander M. Wolff, Phillip C. Aoto, George Meigs, Jody Weng, Taylor Dennis, Michael C. Thompson, Nicholas K. Sauter, Jeff Perry, James Holton, Susan Taylor, Michael E. Wall, and James S. Fraser

Introduction

Diffuse scattering contains a wealth of information about protein motions, and is often present in X-ray diffraction data[1]. Nonetheless, diffuse scattering is rarely used to model and to understand these motions due to the challenge of measuring and interpreting the data[2]. Since the diffuse signal underlies and surrounds the Bragg peaks, it can be challenging to measure both signals well. Initial studies relied upon special detectors to prevent overloaded CCD pixels from obscuring the diffuse signal[3]. Then, it was discovered that pixel array detectors (PADs) enable simultaneous collection of Bragg and diffuse data without special adaption of the experiment[4]. During these experiments, computational removal of background diffuse scatter helped to isolate the signal driven by the crystal. However, as the field advanced it became clear that the optimal measurement of diffuse scattering requires elimination of the background signal during the experiment[5]. Specifically, anisotropic sources of background scatter must be removed at the experimental level. When anisotropic background contaminates the measurements our ability to model protein motions is limited[2]. With these considerations in mind, we set out to discover the experimental parameters that enable collection of the highest quality diffuse scattering signal from protein crystals.

Methods

Sample Preparation

While testing different collection strategies, a host of samples were utilized, including Hen egg-white lysozyme (HEWL), Cyclophilin A (CypA), Htra1, and Protein Kinase A (PKA). HEWL was crystallized using standard conditions, CypA was purified and crystallized as previously described[6,7]. Htra1 crystals were provided by Taylor Dennis and Jeff Perry, with extensive efforts to optimize for larger crystals. PKA crystals were provided by Phillip Aoto, Jody Weng, and Susan Taylor. Phillip extensively explored phase space and seeding methods to optimize for the largest crystals possible, and most diffuse scattering measurements were conducted using PKA crystals.

Diffuse data collection

We collected X-ray data at beamline 8.3.1 at the Advanced Light Source. The beamline was equipped with a Dectris Pilatus3 S 6M detector, and crystals were maintained at controlled temperatures by modulating the flow rate and gas mixture settings on the cryojet. Crystals were looped in Mitegen Microloops™, which are made of Kapton® film that contributes minimal diffuse scatter. To maintain crystal hydration during room-temperature collection, one of the following was applied: 1) Paratone-N oil, by pulling a looped crystal through a droplet of oil; 2) a Mitegen MicroRT™ Capillary, filled at the end with 10 µL of well solution; 3) a Hampton Glass Number 50 (borosilicate glass) capillary, filled at the end with 10 µL of well solution; 4) crystal open to ambient air with Mitegen Watershed™ hydration tip placed within 1 cm of the crystal (the tubing was elongated to 10 feet to enable placement of the tip of the hose within the 8.3.1 mini-hutch), with the Watershed™ set to > 90% relative humidity. To collect background datasets, the goniometer was moved back in the Z-plane so that only the capillary was in the path of the X-ray beam. For data collection, we explored a series of different exposure times, but found best success using 0.1 second exposures with 0.1º oscillations per image. Beam energy was 11.1 keV, with beam divergence set to 0.25 X 0.25 mrad. We used a 100 µm

pinhole to ensure the size of the beam matched the size of the crystal. To minimize radiation damage, we aimed for less than 80 kGy of exposure at each location on the crystal, translating along the crystal once we had exposed 500 images (50º). By adjusting phi at each location, we aimed to collect a minimum of 180º of data for each dataset. When necessary, data was merged across crystals to enhance sampling of the diffuse scattering signal.

Data processing

Bragg data were processed in XDS[8] and Dials[9]. Radiation damage was determined on a per-frame basis using R_d and overall B-factors; images with evidence of radiation damage were culled from both Bragg and diffuse analysis. For diffuse analysis, a combination of the LUNUS[10] software package and custom scripts were utilized. Current versions of both are maintained in the LUNUS GitHub repository (https://github.com/mewall/lunus/). Briefly, experimental geometry was extracted from files output by Dials during Bragg processing. Crystalline datasets and background datasets were then matched by phi angle, and subtraction occurred on a per-image basis. Images were then processed in the LUNUS pipeline. Bragg peaks were removed with a mode filter, intensities were corrected for polarization and normalized as a function of solid-angle, then the data were integrated onto an oversampled reciprocal lattice. The quality of merged diffuse intensities was assessed using tools within LUNUS, namely a resolution-dependent linear correlation coefficient was calculated between symmetrized vs unsymmetrized datasets. Initial models of protein motion were developed using methods within LUNUS, including liquid-like motions (LLM) and rigid body translation (RBT) models. These models were optimized against each dataset, with linear correlation coefficients used to assess the quality of fit.

Results

To begin our investigation into the collection of high-quality diffuse scattering measurements, we carefully measured the background scattering inherent to the experimental setup (**Figure 5.1**). The background measurements revealed a variety of features, including

anisotropic lunes created by the diffuse scatter from Mitegen MicroRT™ Capillaries used to keep crystals hydrated at room temperature. Since anisotropic sources of diffuse scatter are particularly difficult to remove computationally, we asked how we could alter the experimental setup to reduce the background signal, especially sources of anisotropic background signal.

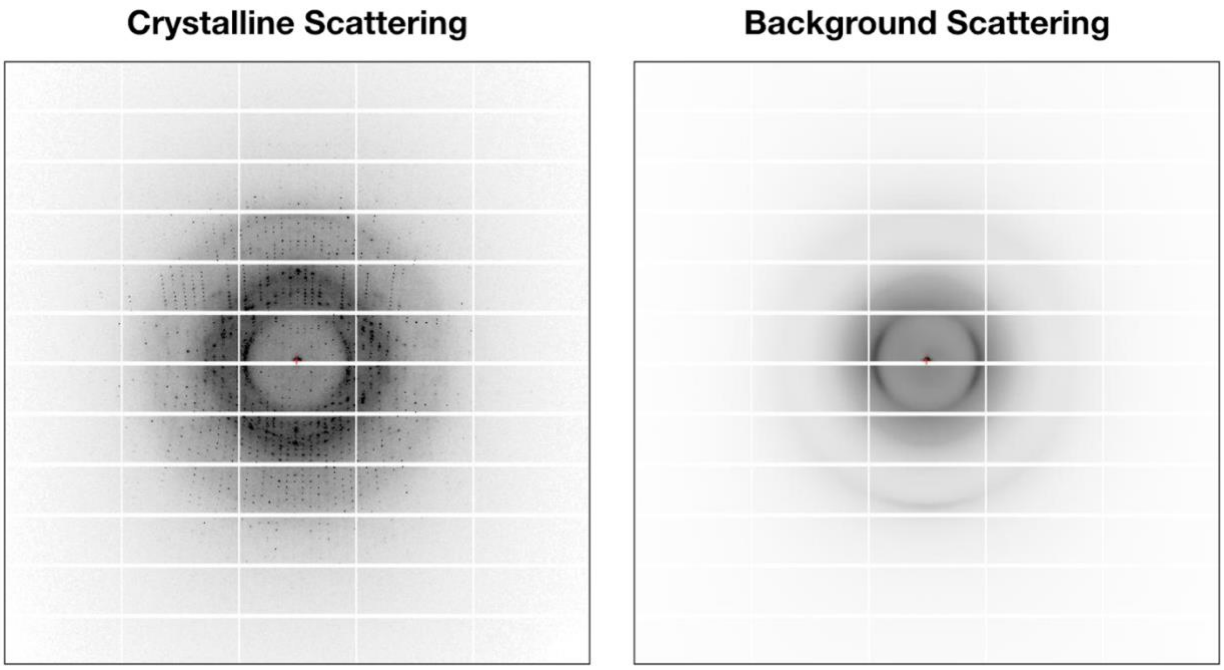

Figure 5.1. Comparison of background and crystalline diffuse scattering. When collecting diffuse scattering from a crystalline sample (left panel), there is an underlying background signal (right panel) that is contributed by the experimental setup.

To systematically remove sources of background scatter, we considered all components of the experimental setup that were external to the crystal (**Figure 5.2**). While not all sources of background scatter could be effectively eliminated, the strongest sources of background scatter included exchangeable items, such as the hydration system. These included Mitegen MicroRT™ Capillaries, Hampton Glass Number 50 (borosilicate glass) capillaries, Paratone-N oil, and innovative use of a Mitegen Watershed™. MicroRT™ capillaries are our standard hydration system, but we also often use Paratone-N oil. Nonetheless, we found that the shape of the Paratone drop varies, leading to anisotropic diffuse scattering that is very difficult to

measure without the crystal in place. Glass capillaries presented another alternative, and tended to scatter very isotropically, making it easy to subtract their contribution via computational methods. On the other hand, we found that data quality was decreased when using glass capillaries, perhaps due to dehydration of the crystal due to poor sealing. Finally, the Watershed™ offered a novel solution, and we were able to minimize background scatter when we used it. To our surprise, Watershed™ data was still inferior to that measured using the MicroRT™ capillaries, despite the reduction in background signal. When the quality of diffuse data was assessed via resolution-dependent correlation coefficients, data collected in Mitegen MicroRT™ capillaries was deemed the highest quality. Similarly, downstream modeling of protein motions was most fruitful when using MicroRT™ datasets, so we decided to proceed with this hydration method.

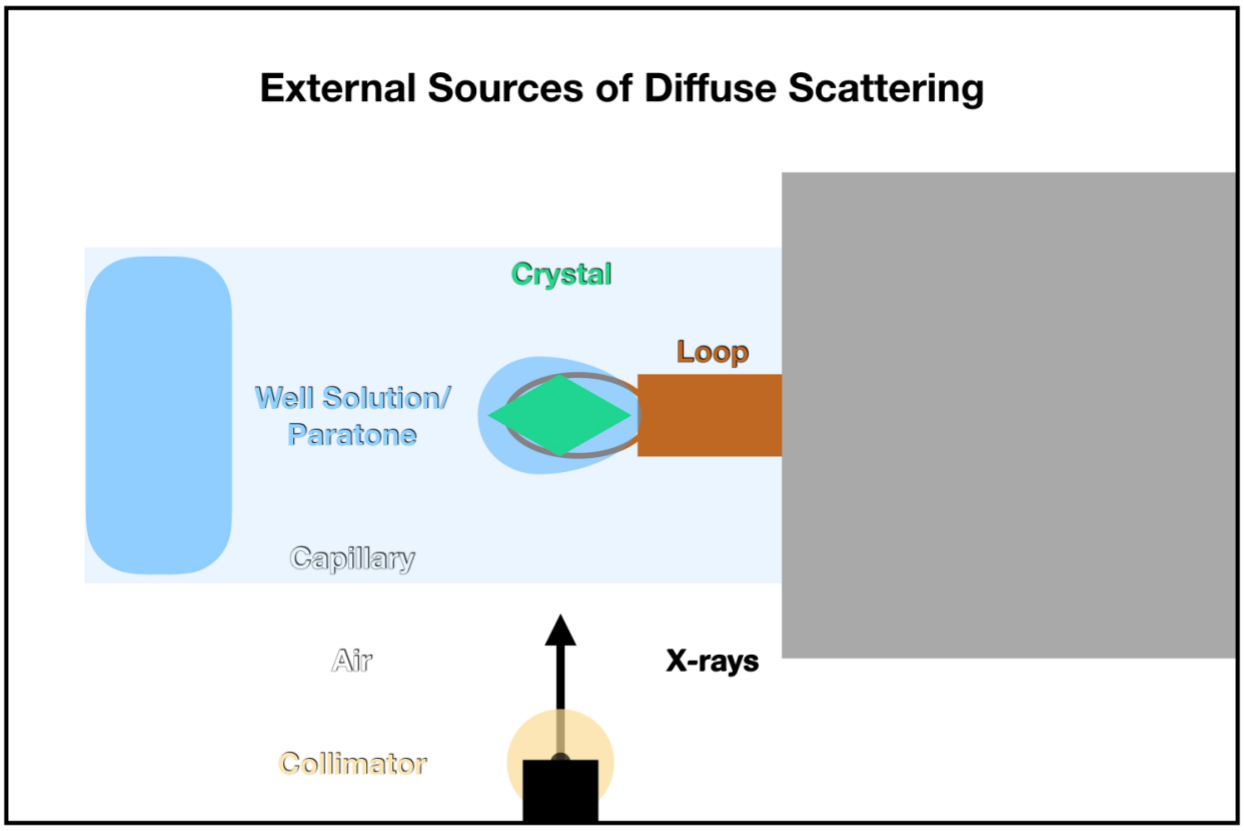

Figure 5.2. External sources of diffuse scattering. Components of a typical room-temperature diffraction setup that contribute to the diffuse scattering signal. Some components

may be altered or removed to minimize the diffuse background that complicates precise measurement of the crystalline diffuse scattering.

A weakness of the Mitegen MicroRT™ Capillaries is that they lead to significant anisotropic background signal (**Figure 5.1**). To overcome this we collected matching background datasets at each phi angle, and subtracted the background signal on a per-image basis (**Figure 5.3**). With this simple method, we were able to minimize the contribution of external sources of diffuse scattering to the diffuse signal. Since our main interest was to investigate protein motions, we then turned our attention to the crystals themselves, asking what factors internal to the crystal might modulate the intensity and the quality of the diffuse signal.

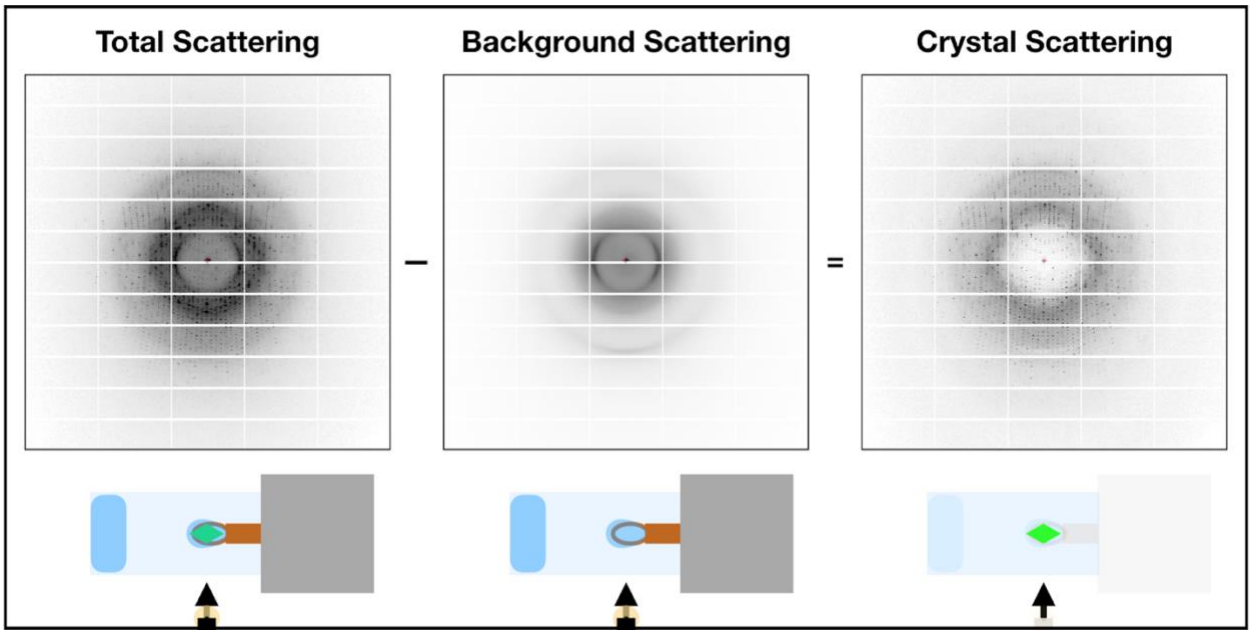

Figure 5.3. Subtraction of diffuse background. By collecting matching background datasets, parasitic sources of diffuse scatter are subtracted from the crystalline dataset, effectively isolating the diffuse scatter that is attributed to the protein crystal.

After minimizing external sources of diffuse scattering, we began to investigate parameters of the experiment that altered the diffuse scatter within crystals. Initially we noticed variations in the mosaicity and the strength of the diffuse scatter from our PKA crystals. To investigate more thoroughly, we zeroed in on one crystal, and collected three sequential datasets from it (**Figure 5.4**). From this single crystal, we saw large variations in mosaicity and

diffuse scatter. Nonetheless, we were unable to determine the cause since we could not isolate a single parameter. For example, we needed to translate along the crystal to avoid accumulating radiation damage, but variations in the local order of the lattice could be driving the differences observed in the data. Furthermore, as we waited for the temperature to equilibrate, the crystal sat for longer periods of time, which may have led to dehydration. Thus, it was difficult to determine which parameter altered the quality of the Bragg and diffuse scattering the most, but since temperature was the most controllable parameter we focused on that. At 284K we observed lower mosaicity and stronger diffuse scatter, so we shot all future PKA datasets at 284K.

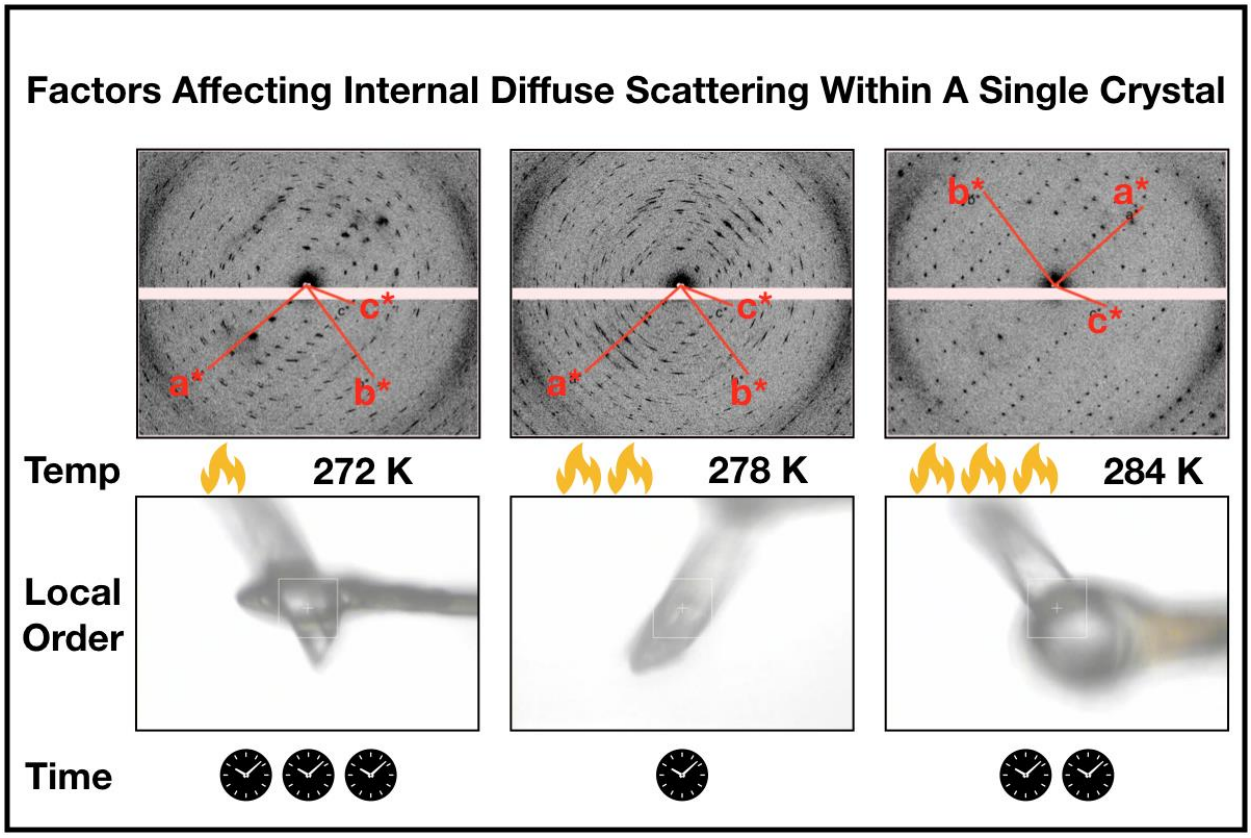

Figure 5.4. Factors affecting internal diffuse scattering within a single crystal. When collecting diffuse scattering data from protein crystals, it is important to consider experimental variables that may modulate the diffuse signal. For example, the temperature may modulate protein conformations, protein dynamics, and lattice disruptions. Local regions of the crystal may be inherently disordered or damaged due to handling. Time from harvest until data collection may affect hydration status, and thus alter the diffuse signal that is observed.

Discussion

While the optimal collection strategy is still an open question, the field of macromolecular diffuse scattering has improved in leaps and bounds over the past few years. From early beginnings using specialized detectors[3], to simultaneous measurement of Bragg and diffuse scatter using PAD detectors[4]. Recent innovations include the first XFEL diffuse scattering dataset[11], as well as dramatic improvements in data quality that enabled clear insights into modeling protein motions[5]. While our work to eliminate background scattering led to fruitful experiments, we ultimately converged with Meisburger *et al.*[5] on a simple frame-by-frame subtraction strategy. This is not to say that future innovators may not improve the measurement of diffuse scattering by using creative methods such as the Mitegen Watershed™, but for now the highest quality datasets have been collected under more traditional hydration schemes. Finally, we look forward to future work investigating the effects of temperature and hydration status upon diffuse scattering from protein crystals, and hope that the next five years are as fruitful as the last.

References

1. Meisburger, S. P., Thomas, W. C., Watkins, M. B. & Ando, N. X-ray Scattering Studies of Protein Structural Dynamics. *Chem. Rev.* **117**, 7615–7672 (2017).
2. Wall, M. E., Wolff, A. M. & Fraser, J. S. Bringing diffuse X-ray scattering into focus. *Curr. Opin. Struct. Biol.* **50**, 109–116 (2018).
3. Wall, M. E., Ealick, S. E. & Gruner, S. M. Three-dimensional diffuse x-ray scattering from crystals of Staphylococcal nuclease. *Proc. Natl. Acad. Sci.* **94**, 6180–6184 (1997).
4. Benschoten, A. H. V. *et al.* Measuring and modeling diffuse scattering in protein X-ray crystallography. *Proc. Natl. Acad. Sci.* **113**, 4069–4074 (2016).
5. Meisburger, S. P., Case, D. A. & Ando, N. Diffuse X-ray scattering from correlated motions in a protein crystal. *Nat. Commun.* **11**, 1–13 (2020).
6. Keedy, D. A. *et al.* Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife* **4**, e07574 (2015).
7. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
8. Kabsch, W. *XDS*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
9. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 85–97 (2018).
10. Wall, M. E. Methods and Software for Diffuse X-Ray Scattering from Protein Crystals. in *Micro and Nano Technologies in Bioanalysis: Methods and Protocols* (eds. Foote, R. S. & Lee, J. W.) 269–279 (Humana Press, 2009). doi:10.1007/978-1-59745-483-4_17.
11. Ayyer, K. *et al.* Macromolecular diffractive imaging using imperfect crystals. *Nature* **530**, 202–206 (2016).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
40CC14E058D14F5...
Author Signature

3/16/2020
Date